Universidad de las Ciencias Informáticas

Facultad 1

Título: Propuesta de filtrado de correo electrónico para el sistema Smart Keeper

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

Autor: Indira Escalona Martínez

Tutores: Ing. Dovier Antonio Ripoll Méndez

Ing. Yaneida Rondón Hernández

La Habana, Cuba

Junio de 2012

"Año 54 de la Revolución"

Pensamiento

"La soberanía del hombre está oculta en la dimensión de sus conocimientos."

Sir Francis Bacon

Declaración de autoría

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

Para que así conste firmamos la presente a los ___ días del mes de ___ del año 2012.

_____

Indira Escalona Martínez

_____ _____

Ing. Dovier Antonio Ripoll Méndez Ing. Yaneida Rondón Hernández

Agradecimientos

A a todos los que, de una forma u otra, me han ayudado en el desarrollo de este Trabajo de Diploma, esta tesis también es suya.

A todas las personas maravillosas que he conocido en estos 6 años.

Gracias

Dedicatoria

A mis padres y mis abuelos.

A la familia en general.

Resumen

En la Universidad de las Ciencias Informáticas se desarrolla, desde el año 2005, el filtro de contenido web Smart Keeper. Su principal objetivo es regular (permitir o denegar) el acceso a contenidos inadecuados según las reglas establecidas en las instituciones donde sea instalado. Actualmente dicho sistema no cuenta con una herramienta que realice el filtrado de correo electrónico. Smart Keeper posee funcionalidades que pueden ser aprovechadas en el filtrado de correo electrónico.

El objetivo de este trabajo consistió en elaborar una propuesta de filtrado de correo electrónico que hiciera uso de las herramientas ya implementadas en Smart Keeper. Para lograr dicho objetivo ha realizado un estudio acerca de las herramientas que realizan el filtrado de correo, así como las técnicas que estos utilizan para detectar los correos no deseados y los virus.

En la elaboración de dicha propuesta se han utilizado herramientas como los módulos de MOCIC y MOCICE para lograr mayor independencia en la categorización de los mensajes. También, se utilizó el programa SpamAssassin para realizar el análisis anti *spam* y las listas negras/blancas de direcciones. Además de utilizar los módulos existentes en Smart Keeper. Para validar dicha propuesta se han utilizado la opinión de expertos que permitió asegurar la viabilidad de esta.

Palabras claves: filtrado, correo electrónico, Smart Keeper, spam, virus, filtro, módulos (1).

Índice

Introducción

El desarrollo de las tecnologías digitales ha provocado la expansión progresiva de Internet. Se estima que en el período 2000-2011, en América Latina hubo un incremento de los usuarios en un 1037.4 % y en África de un 2527.4 % (1), lo cual muestra su crecimiento en las regiones menos desarrolladas, como puede apreciarse en la Tabla 1. Esto ha traído consigo un aumento de los servicios que se brindan en la red de redes. Entre los servicios más populares se encuentran la mensajería instantánea, las redes punto a punto, la transmisión de archivos, la telefonía Voz IP, el acceso remoto a otros dispositivos y el correo electrónico.

El correo electrónico constituye uno de los servicios más extendidos y utilizados para la comunicación electrónica (2). Ofrece muchas ventajas, como son su alta disponibilidad, su accesibilidad y su rapidez en el recibo/envío de un mensaje. En el año 2011 un usuario corporativo promedio envió y recibió al menos 112 mensajes al día (3), evidenciándose las características antes mencionadas. Otra de sus ventajas es el envío de mensajes a múltiples destinatarios, evitando la pérdida de tiempo y recursos en hacer copias de un correo. Mediante un correo también es posible enviar todo tipo de archivo digital, como por ejemplo videos, ficheros, imágenes y gráficos, enriqueciendo la comunicación. Lo anterior ha motivado que en casi todo el mundo, existan más de 1.88 billones de usuarios con cuentas de correo. La popularidad de este servicio ha motivado la publicación de disímiles estudios y artículos que tratan temas relacionados con sus ventajas y desventajas. Algunos de ellos señalan el tiempo de estudio/trabajo perdidos por los usuarios debido al uso excesivo o no moderado de este servicio, provocando que el 71 % del tráfico mundial de correo sea *spam* o correo basura como también se le conoce (2). Otros de los inconvenientes que puede provocar esta tecnología son:

- La obtención de información personal o confidencial a través del correo.
- El recibo de correos no deseados con información falsa o publicidad.

- Los adjuntos pueden contener virus o contenidos no permitidos.
- Los mensajes pueden ser de origen *spam* para provocar la caída de los servicios de la red.

Todas las desventajas o inconvenientes antes mencionados han propiciado la aparición de los filtros de correo electrónico. Estos sistemas implementan un grupo de técnicas de filtrado con el objetivo de bloquear y/o detectar los correos no deseados o infectados por virus. Algunas de estas técnicas pueden ser de tipo anti *spam* o antivirus. Dichas técnicas se consolidan en programas que pueden formar parte de filtros de correo electrónico, o pueden constituir soluciones de filtrado independientes.

En el caso de los programas anti *spam* la detección puede ser manual, en la cual una comunidad de usuarios comparte información acerca de las fuentes de *spam*. Es importante destacar que ningún anti *spam* es infalible debido a que siempre hay un por ciento de los correos filtrados que resultan ser falsos negativos[1] o pueden ser categorizados como correos *spam* sin serlos.

Los programas antivirus actualmente incorporan la funcionalidad de desinfectar, poner en cuarentena o eliminar el archivo. El progresivo desarrollo de las tecnologías ha provocado la aparición de virus más fuertes contra los cuales han surgido soluciones cada vez más eficaces haciendo que los software añadan mejores funcionalidades y haya una amplia variedad de soluciones a disposición de los usuarios.

Además de poner en práctica técnicas y métodos para detectar correos *spam* o que estén infectados con virus, los filtros de correo electrónico implementan otras funcionalidades en dependencia de necesidades específicas de los usuarios. Ejemplo de ello, estos pueden permitir el bloqueo de puertos, la encriptación de los mensajes o el filtrado en varios idiomas.

[1] Un mensaje que realmente es spam (correo no deseado), pero que no ha sido catalogado como tal por el sistema.

Los filtros de contenido web constituyen una solución a la problemática de regular el acceso de los usuarios al contenido presente en Internet. Esta necesidad viene dada por la presencia de materiales tanto educativos, científicos, culturales como nocivos, ilícitos o incluso ilegales para algunos países en dependencia de sus legislaciones.

En la Universidad de las Ciencias Informáticas (UCI) desde el año 2005 se desarrolla el filtro de contenido Smart Keeper[2]. Este sistema basa su filtrado en reglas o políticas establecidas para la navegación de los usuarios y en Uniform Resource Locators[3] (URLs) preclasificadas tomadas de los sitios proveedores Issak y Toulouse[4]. Dichas políticas agrupan elementos como la restricción de ficheros, el análisis antivirus, la utilización de expresiones regulares, entre otros.

El sistema, en vías de lograr una mayor independencia en la categorización de los contenidos de Internet, será integrado en próximas versiones con el Motor de Clasificación Inteligente de Contenido (MOCIC) y el Motor de Clasificación Inteligente de Contenido de Correo Electrónico (MOCICE). Estos motores realizan la clasificación de contenidos a partir de algoritmos de inteligencia artificial haciendo uso de categorías preestablecidas.

Actualmente Smart Keeper solo realiza su filtrado sobre los contenidos de la web. Una vez desplegado, el sistema se limita a regular el acceso a los materiales en Internet, sin extenderse al tratamiento de otros

2 Antes conocido como Filpacon

3 Traducción de la autora: Localizadores de recursos uniforme

4 Sitio oficial del proveedor: http://dsi.ut-capitole.fr/blacklists/

servicios como lo es el correo electrónico. Para este fin se pudieran utilizar además las características mencionadas de dicho sistema de filtrado, sin embargo, no existe una solución que las aproveche.

Persiguiendo que el sistema Smart Keeper sea una solución más completa y que pueda suplir todas las necesidades de las empresas o instituciones que lo utilicen, se hace necesario que dicho sistema se provea de un mecanismo para realizar el filtrado de correo electrónico.

A partir de la problemática antes descrita surge el siguiente **problema investigativo:** ¿Cómo realizar el filtrado de correo electrónico en el sistema Smart Keeper?

Se define como **objeto de estudio:** el filtrado de servicios de Internet y el **campo de acción:** el filtrado de correo electrónico.

Como **objetivo general:** Elaborar una propuesta de filtrado de correo electrónico en el sistema Smart Keeper para ampliar el alcance de dicho producto respecto al tratamiento de servicios de Internet. Del cual se derivan los siguientes **objetivos específicos:**

- Profundizar acerca del filtrado de correo electrónico.
- Describir la propuesta de filtrado de correo electrónico en el sistema Smart Keeper.

Se establece como **idea a defender:** La elaboración de una propuesta de filtrado de correo electrónico en el sistema Smart Keeper permitirá ampliar el alcance de dicho producto respecto al tratamiento de servicios de Internet.

En el cumplimiento de la tarea se usarán los siguientes **métodos teóricos:**

Analítico-Sintético:

Facilitará el entendimiento del tema de investigación (filtrado de correo electrónico), realizando una división de este en diferentes fases con el fin de definir sus principales características. Permitirá el entendimiento de cada uno de los elementos que componen el correo electrónico y el correspondiente filtrado de este; para luego establecer las relaciones y características generales que ambos procesos poseen.

Histórico-Lógico:

Este método trazará una trayectoria del objeto de estudio para mostrar las diferentes y principales etapas por las que ha transitado, a la vez que decantará la esencia de este estudio. De forma que se muestren los precedentes del filtrado de correo electrónico y las condiciones que propiciaron su surgimiento.

Estructura del informe:

Capítulo 1: Este capítulo recogerá la fundamentación teórica del tema. Se presentará el estado del arte mediante el estudio de soluciones a nivel nacional e internacional. Además, se abordarán temas relacionados con el filtrado del correo electrónico. Se realizará un resumen de las tendencias y los métodos que se utilizan en el filtrado de dicho servicio.

Capítulo 2:

En este capítulo se instrumentará una descripción de la propuesta que pretende este trabajo, así como la validación de la misma.

Capítulo 1. Filtrado de correo electrónico

Introducción

Los filtros de correo electrónico son herramientas que permiten detectar o bloquear la entrada de correos no deseados o infectados con virus a los buzones de los usuarios. La implementación de esta funcionalidad requiere el estudio de todos los aspectos relacionados a ellos con el objetivo de entender mejor su funcionamiento. En este capítulo se abordarán todos los términos relacionados a este tema, las técnicas utilizadas en el filtrado de correo electrónico así como algunos de los sistemas que actualmente las implementan.

1.1 Trabajos similares

En la actualidad existen un conjunto de aplicaciones que desarrollan técnicas para detectar los correos *spam* y virus. A continuación se realizará un estudio de los sistemas de filtrado de correo electrónico para identificar los métodos más apropiados para dichos procesos así como conocer las características de las soluciones actuales. Primero se hará un estudio general acerca de los software de filtrado de correo electrónico para luego profundizar sobre los filtros de anti *spam* y antivirus como productos independientes. Dicha investigación sentará las bases para el desarrollo de una propuesta que se adecúe mejor a las necesidades del sistema Smart Keeper.

1.1.1 Filtros de correo electrónico

Astaro Mail Security

Astaro AG es un proveedor de soluciones de seguridad. Hoy día la compañía desarrolla y comercializa sus soluciones de seguridad tanto de hardware como de software. Entre sus productos se encuentra el

Astaro Security Gateway Software Appliance dentro del cual se incluye el Astaro Mail Security, un componente cuyo objetivo es el filtrado de correo electrónico.

El motor principal de esta herramienta utiliza para la detección de correos *spam*, los patrones globales y las tecnologías de huellas digitales, no tiene en cuenta el idioma del correo o el contenido de este. Los correos son identificados mediante la comunicación en tiempo real con una red de analizadores y participantes que escanean electrónicamente y comparten las huellas digitales de los correos que ellos reciben. Además, este sistema de correo electrónico identifica los correos no solicitados. Incluso provee la funcionalidad de cifrado de correo electrónico en un dispositivo de Gestión Unificada de Amenazas. Los correos electrónicos y archivos adjuntos pueden ser rechazados con un mensaje al remitente, o aprobada con una advertencia o en cuarentena.

Las tecnologías de antivirus de este software detectan y bloquean virus, troyanos, gusanos y otros software maliciosos que se encuentran en los correos electrónicos, la Web y las descargas de ficheros a través del Protocolo de Transferencia de Fichero, en inglés *File Transfer Protocol* (FTP), antes de que ellos lleguen a los servidores de correo o los escritorios.

Panda Cloud Email Protection

Dentro del catálogo de productos de la empresa PANDA[5] se encuentra este software que realiza el filtrado de correo spam, reduciendo el consumo de recursos y ancho de banda de la empresa. La cuarentena de Panda Cloud Email Protection almacena todo el correo con *spam* o que contenga virus potenciales. Esto asegura que los servidores de los clientes no se saturen de *spam*, en consecuencia, se liberan recursos y

[5] Empresa desarrolladora de los software del mismo nombre.

capacidad. Además de reducir el índice de virus que entra a la red. Reduce el uso del ancho de banda y los procesos del servidor.

Ofrece un conjunto de opciones para el manejo de los correos no deseados. Permite bloquear todo el tráfico no deseado dentro de los perímetros de la red. También, posee una interfaz web para la administración del sistema. Debido a que este producto es privativo no se puede acceder a más detalles de su implementación y funcionamiento.

Email ThreatPak

En este producto cada correo electrónico se analiza y se le asigna un puntaje de spam por las diversas tecnologías (Bayesiano, heurística, listas negras, listas blancas, la reputación) incrustadas en el motor SpamFilter Spam; estos pueden ser rechazados, puestos en cuarentena o entregarse normalmente, en función de su puntuación de *spam* y la configuración de umbral. Este software de filtrado escanea todos los correos electrónicos entrantes y salientes con las palabras clave definidas por el usuario y frases encontradas en el cuerpo del correo electrónico, así como en muchos tipos de archivos adjuntos. Este filtro de correo electrónico posee una amplia base de datos de más de 300 000 amenazas de *malware*. Además cuenta con un sistema de reportes que mantiene informado al administrador de dicho sistema acerca de la actividad en el correo.

Los sistemas de filtrado de correo antes analizados además de ser privativos no permiten una completa integración con el sistema Smart Keeper lo cual imposibilita aprovechar las funcionalidades ya implementadas en él. Aun así, el análisis de las herramientas antes mencionadas permitió profundizar acerca de los sistemas anti *spam* y antivirus. Es necesario hacer un estudio de las técnicas identificadas con el objetivo de implementar un filtro de correo electrónico que haga uso de los datos almacenados en la base datos de Smart Keeper para realizar filtrado eficiente.

1.1.2 Filtros anti spam

Los filtros anti *spam* por si solos no constituyen una solución de filtrado de correo electrónico; pero desempeñan un papel importante en el proceso de filtrado debido a que uno de los principales objetivos de estos sistemas es la detección y bloqueo de los correos no deseados. A continuación se presentan varios ejemplos de ellos para el análisis de sus características:

SPAMfighter Mail Gateway[6]

Este es un software de filtrado considerado uno de los mejores filtros actualmente en el mercado (4). Los sistemas operativos que se requieren para su funcionamiento son Windows 2000 Server, Windows 2003 Server, Windows 2008 o las últimas versiones de este, ya sean las versiones de 32 o 64 bits.

Tiene a su disposición un grupo de cerca de 8 millones de personas para el reporte de correos *spam* lo que proporciona un bajo índice de falsos positivos. Utiliza las listas negras de dominios, el filtrado por idiomas, las listas blancas y negras de direcciones e implementa algoritmos de aprendizaje para la detección de los correos indeseados. Su actualización se realiza automáticamente. Los archivos adjuntos son escaneados y en caso de estar infectados son eliminados.

Existen otras soluciones de la misma empresa con diferentes características entre los que se puede encontrar el SPAMfighter Estándar que es gratis (5).

Cloudmark DesktopOne Pro Mode[7]

---

[6] Sitio oficial del producto: http://www.spamfighter.com

[7] Sitio oficial del producto: http://www.cloudmark.com

También es uno de los mejores filtros de spam encontrados actualmente. Permite escanear carpetas de correos automáticamente y seleccionar el momento en que se va a procesar cada carpeta. Maneja automáticamente varias cuentas de correo y facilita el filtrado de cuentas de correo POP, IMAP, Exchange y Web.

Además, brinda la posibilidad de consultar estadísticas relacionadas a las cuentas de correo que son protegidas por el filtro. Es de fácil instalación y una vez instalado el programa este detecta automáticamente las cuentas de correo. La versión libre de este producto se nombra DesktopOne Basic Mode.

MailWasher Enterprise Server [8]

MailWasher Enterprise Server es un software de filtrado de spam que trabaja sobre todas las versiones de servidores de correo de Windows y MTAs de GNU/Linux como Sendmail, Postfix, QMail, entre otros. Permite borrar los correos en el servidor antes de descargarlos a la máquina. Implementa técnicas para la detección de los correos no deseados que van desde las listas de direcciones permitidas, las listas negras personales, filtrado bayesiano hasta listas negras externas. Se instala y configura con facilidad. Puede procesar 4000 mensajes por minuto utilizando pocos recursos de la máquina, lo cual disminuye la carga del servidor. El aprendizaje inteligente y la actualización de las firmas de *spam* se realizan en tiempo real.

SpamAssassin[9]

SpamAssassin es un proyecto de *Apache Software Fundation* liberado bajo la licencia Apache. Aplica una variedad de pruebas tanto al contenido del correo como a los encabezados usando métodos estadísticos

8 Sitio oficial del producto: http://www.mailwasher.net/

9 Sitio oficial del producto: http://spamassassin.apache.org/

avanzados. Este programa en Perl utiliza la puntuación combinada de los distintos controles que posee para determinar si es un correo *spam*.

Sus características son:

- Análisis de cabecera y cuerpo del mensaje de forma independiente.
- Reglas de búsqueda de texto común en los correos de *spam*.
- Filtros bayesianos de autoaprendizaje.
- Pruebas de red para comprobación de fuentes de *spam* y spammers.
- Listas negras/blancas de *spam* configurables.
- SURBLs, son dominios contenidos en correos *spam*, e indicadores de fuentes de *spam*.
- Las pruebas anti *spam* y de configuración se almacenan en texto sin formato, por lo que es fácil de configurar y añadir nuevas reglas.

Ventajas (Ozonosecurity[10])

- SpamAssassin utiliza diferentes métodos para calcular si un correo electrónico es *spam*, a diferencia de otras tecnologías que usan solo un método.
- Es altamente configurable, posicionándolo como un sistema flexible que se adapta a necesidades específicas.

[10] Sitio donde se evalua el producto SpamAssassin: http://www.ozonosecurity.com

- Es adaptable a nuevas técnicas de los “spammers”. A través del desarrollo de nuevos *plug-ins*, es sencillo añadirle nuevos métodos o funcionalidades.
- Existe una comunidad de desarrolladores buscando siempre nuevas maneras de mejorar la tecnología y sus métodos de detección de *spam*.
- SpamAssassin contiene un módulo de auto aprendizaje, su filtro bayesiano, que examina el *spam* y se adapta automáticamente a las nuevas técnicas desarrolladas por los spammers.

Debe destacarse que este es un programa de software libre y permite su utilización en otras soluciones comercializables.

Entre todos los sistemas anti *spam* analizados, se ha seleccionado para la detección de mensajes *spam* el programa SpamAssassin debido a las ventajas antes mencionadas y a la amplia documentación que este posee. Además, facilita la integración a otros programas existentes lo cual lo hace muy flexible a otras configuraciones. Aun así este software no permite hacer uso de elementos como las expresiones regulares, las políticas y el análisis antivirus de Smart Keeper; haciendo necesario la búsqueda de otras técnicas que puedan hacer uso de los datos de Smart Keeper.

1.1.3 Antivirus

Los programas antivirus actualmente implementan sistemas que detectan correos *spam* y virus a la vez, ofreciendo un amplio conjunto de configuraciones y opciones para su eliminación o desinfección. Actualmente Smart Keeper realiza análisis antivirus en su filtrado, siendo esta una característica que puede ser aprovechada para el filtrado de correo electrónico. La versión actual de dicho sistema solo

soporta la utilización del antivirus Clamav, por tanto no es objetivo de este trabajo proponer la utilización de antivirus similares. A continuación se muestran características y ventajas del sistema Clamav.

Clamav Antivirus[11]

Es un antivirus de software libre con Licencia GPL que surge con el principal objetivo de combatir los correos maliciosos, por lo cual actualmente está instalado en varios servidores de correo electrónico. Actualmente usan este antivirus Apple, Linux.org, IBM, OnLAMP, Institut Pasteur, SourceForge, Universidades en Italia, Brasil, USA, España, pequeños y medianos negocios. Muestra de su uso extendido es en la Loyola University of Chicago donde tienen cerca de 30 000 usuarios (7).

Sus principales **características** son:

- POSIX compatible y portable.
- Escaneo rápido.
- Detecta virus, gusanos y troyanos, incluyendo virus programados como macros de Microsoft Office. La base de datos se actualiza varias veces al día.
- Escaneo de archivos y ficheros comprimidos: ZIP, RAR, ARJ, TAR, Gzip, Bzip2, MS OLE2, MS Cabinet File, MS CHM, MS SZDD, BinHex, SIS, y otros.
- Soporta plataformas de 32/64 bit.
- Soporta la mayoría de formatos de correo electrónico.

[11] Sitio oficial de la aplicación: http://www.clamav.net

- Soporta formatos especiales como: HTML, RTF, PDF y los ficheros de Microsoft Office y MacOffice.

Ventajas

- Solución madura y estable para servidores de Mail, Web, Proxy, repositorios.
- Solución libre.
- Acceso al código fuente.
- Comunidad internacional con alta capacidad de reacción.

En el ámbito nacional no se tiene conocimiento de ninguna herramienta para el filtrado de correo electrónico. Aun cuando se han encontrado herramientas que pueden facilitar la detección de correos no deseados o virus; no se conoce de una solución para filtrar el correo electrónico que haga uso de las características del sistema Smart Keeper. Por tanto se hace necesario proponer una solución a la medida, que supla dicha necesidad.

1.2 Funcionamiento de Internet

Internet, o la red de redes, surgió a partir de un proyecto estrictamente militar llamado ARPANET, en 1969. Esta posteriormente fue incrementando sus usuarios con la conexión a ella de universidades norteamericanas y europeas, mayoritariamente para estudios con fines científicos. Según el diccionario de la Lengua Española de la Real Academia la Internet es una *Red informática mundial, descentralizada, formada por la conexión directa entre computadoras u ordenadores mediante un protocolo especial de comunicación.* (8)

El funcionamiento de internet se basa en tres **elementos fundamentales** (9):

- Los **protocolos de comunicación**: son un conjunto de normas que se establecen de forma que todos los ordenadores en los distintos lugares del mundo puedan intercambiar datos. Con el crecimiento de Internet y el objetivo de que esta fuera una red mundial se acordó la unificación de todos los protocolos, utilizándose actualmente el conjunto *Transmission Control Protocol/Internet Protocol* (TCP/IP). El protocolo IP es la forma en que se van a comunicar las computadoras entre sí, donde el intercambio se realiza a través de paquetes o datagramas de información. El protocolo TCP garantiza que los datos serán entregados en su destino sin errores y en el mismo orden en que se transmitieron. También proporciona un mecanismo para distinguir distintas aplicaciones dentro de una misma máquina, a través del concepto de puerto.
- La **dirección IP**: esta es una dirección única asignada a cada máquina que se conecta a la red. Esta posee 4 grupos de números que pueden tomar un valor entre 0 y 255, aunque no todos los valores están disponibles para designar una dirección IP de usuario válida, debido a que muchos de ellos están reservados para direcciones concretas.
- Los **servidores**: son máquinas que prestan algún tipo de servicio a los usuarios que establecen conexión con ellos, ya sea de correo electrónico, transferencia de archivos, conversación, etc. Estas procesan las peticiones de los usuarios y generan una respuesta adecuada.

Sistemas de nombres de dominio

Los *Domain Name Systems*, en español Sistema de nombres de dominio, es una base de datos distribuida y jerárquica que almacena información asociada a nombres de dominio en redes como Internet. Como base de datos el DNS es capaz de asociar diferentes tipos de información a cada nombre, pero los usos más comunes son la asignación de nombres de dominio a direcciones IP y la localización de los servidores de correo electrónico de cada dominio.

Un Servidor de DNS es una máquina que contiene una base de datos, por lo general cada proveedor de

Internet, *Internet Provider* (ISP) tiene su Servidor de DNS asociado, y aunque es conveniente utilizar este, se puede utilizar cualquier otro. En un Servidor DNS existe una entrada que indica el Servidor de Correo del Dominio, estos servidores son indicados mediante el registro *Mail eXchange* (MX).

Entre los beneficios de Internet se puede mencionar el fácil acceso a la información. Con el uso de las herramientas adecuadas se pueden realizar una infinidad de actividades; constituye una fuente inagotable de recursos educativos y actualidad informativa los cuales pueden ayudar en el enriquecimiento profesional y cultural. La creación de comunidades virtuales ha permitido que personas con interés en un tema en específico puedan intercambiar opiniones. Ha facilitado la obtención de soporte técnico a herramientas, software o hardware. A diario las empresas ofrecen un sinnúmero de servicios a los cuales se puede acceder en segundos.

Si bien tiene muchas ventajas, estas llevan aparejadas varios perjuicios, como son la inseguridad de la información compartida o transmitida por la red ante la tendencia en los últimos años a realizar transacciones a través de ella; la presencia de personas dedicadas a romper las barreras de seguridad de las empresas. La publicación de muchos contenidos, algunos ilícitos y otros no tanto pero que constituyen formas de distracción y consumen el tiempo de los usuarios sin aportar beneficio alguno.

Entre los servicios más utilizados se encuentra la *World Wide Web*, o simplemente la Web, muchas veces confundida con el término Internet que permite intercambiar archivos mediante el Protocolo de Transmisión de Hipertexto, en inglés *Hypertext Transport Protocol* (HTTP) de forma muy sencilla. Ejemplos de otros servicios que se brindan en la red son: la transmisión de archivos, conversaciones en línea, la mensajería instantánea, el acceso remoto a otros dispositivos y el envío de correo electrónico (11).

Actualmente Internet posee más de mil millones de usuarios, según las estadísticas de *Internet World Stats*[12]. Un estudio realizado por la comunidad Red Innova[13] muestra como resultado que el 49.45 % de los usuarios de Internet pasan más de ocho horas conectados. Una amplia mayoría, el 90.76 % para consultar el correo electrónico, siendo este su uso mayoritario (10).

1.3 Correo electrónico

Conocido también como e-mail, del inglés *electronic mail*, se comenzó a utilizar en 1965 en una supercomputadora de tiempo compartido y, para 1966, se había extendido rápidamente para utilizarse en las redes de computadoras. Hoy día se ha convertido en una parte esencial de la vida de millones de usuarios alrededor del mundo. El correo electrónico es una herramienta muy utilizada y extendida para la comunicación electrónica debido a la rapidez, simplicidad y el bajo costo con el cual se puede enviar un mensaje.

Este servicio de red permite a los usuarios enviar y recibir mensajes y archivos rápidamente mediante sistemas de comunicación electrónicos, ejemplo de ello el Protocolo simple de transferencia de correo, en inglés *Simple Mail Transfer Protocol* (SMTP), en conjunto con otros protocolos que lo complementan como son el POP3 e IMAP.

Se debe aclarar que existen dos tipos de correo electrónico y aun cuando lo proveedores de correo suelen brindar el servicio de ambas formas, estas difieren en algunos aspectos.

Tipos de correos (12)**:**

12 http://www.internetworldstats.com

13 http://www.redinnova.com/

Correo web: es el correo que se revisa a partir de una página web mientras que los mensajes siempre van a estar almacenados en el servidor. Por tanto, es necesario un navegador web y solo se va a tener acceso a las funcionalidades que este brinde. Por ejemplo: Gmail, Yahoo y Hotmail.

Correo "POP": requiere que se instale y utilice un software cliente de correo. Los mensajes se quedarán almacenados en el disco duro de la máquina e incorporan más funcionalidades que el correo web. Ejemplos de ellos son: Microsoft Office Outlook, Mozilla Thunderbird y Eudora.

1.3.1 Protocolos de correo electrónico

El **protocolo SMTP** es el protocolo estándar que permite la transferencia de correo de un servidor a otro mediante una conexión punto a punto. Este se basa en el modelo cliente-servidor, funcionando con comandos de textos enviados al servidor SMTP específicamente al puerto 25, a cada comando enviado por el cliente le sigue una respuesta del servidor SMTP compuesta por un número y un mensaje descriptivo. La comunicación entre el cliente y el servidor consiste enteramente en líneas de texto compuestas por caracteres ASCII. El tamaño máximo permitido para estas líneas es de 1000 caracteres. Prácticamente todos los mensajes de correo electrónico escritos por personas en Internet y una proporción considerable de estos mensajes generados automáticamente son transmitidos en formato MIME a través de SMTP (13).

MIME, extensiones multipropósito de correo de internet, son una serie de convenciones o especificaciones para el intercambio a través de Internet de todo tipo de archivos, texto, audio, vídeo, etc. de forma transparente para el usuario. Esta norma fue definida por la IETF, desde 1994 todas las extensiones MIME están especificadas de forma detallada en diversos documentos oficiales disponibles en Internet. Estas flexibilizan el adjuntado de archivos de distintos tipos, así como la inclusión de otros juegos de caracteres que no sean los US-ASCII (caracteres ASCII norteamericanos) como se especificaron en SMTP. Ellas posibilitan el envío de múltiples adjuntos en un solo mensaje, representan una identificación única para cada segmento del mensaje y proporcionan información adicional sobre el contenido del mensaje (13).

El **protocolo POP3** (Protocolo de oficina de correos, *Post Office Protocol*) permite recoger el correo electrónico en un servidor remoto (servidor POP), los correos electrónicos recibidos pueden ser consultados sin estar conectados a internet. Este protocolo fue diseñado para trabajar conjuntamente con el protocolo TCP. Inicialmente el proceso trabaja a través del puerto 110, a la espera de una conexión, una vez y se establece la comunicación comienza un intercambio de comandos y respuestas hasta que la conexión se libera. Los estados del POP3 son: autorización, en el que se entra cuando se establece la conexión TCP y sirve para que los usuarios se identifiquen ante el protocolo; transacción cuando se hace una identificación positiva del usuario que quiere ingresar, aquí los mensajes pasan del servidor al cliente. Una vez finalizado este proceso, se pasa al estado actualización, donde se elimina los mensajes que el usuario recibió, así finaliza la conexión y se libera (13).

El **Protocolo de acceso a mensajes de Internet**, en inglés *Internet Message Access Protocol* **(IMAP)**, es un protocolo alternativo al POP3, y al igual que este es un protocolo de red de acceso a mensajes electrónicos almacenados en un servidor. Este posee varias ventajas sobre el protocolo POP3, permite administrar diversos accesos de manera simultánea a un mismo buzón, administrar diversas bandejas de entrada. En el caso de IMAP4, este proporciona mecanismos para hacer peticiones por parte del cliente al servidor para realizar búsquedas de mensajes de acuerdo a una cierta variedad de criterios, lo que evita que los clientes descarguen todos los mensajes de su buzón de correo, agilizando las búsquedas. También permite visualizar los mensajes de manera remota y no descargando los mensajes. Se conecta al servidor por el puerto 143 (TCP) (13).

1.3.2 Servidores de correo

Un servidor de correo es una aplicación informática que tiene como objetivo enviar, recibir y gestionar mensajes a través de las redes de transmisión de datos existentes, con el fin de que los usuarios puedan mantenerse comunicados con una velocidad mayor. Este software está diseñado en función de protocolos estándares para manejar los mensajes de email, los gráficos que pueda contener y archivos adjuntos. Para realizar esta tarea se utilizan los Agentes de Transferencia de Correo, en inglés *Mail Transport Agent*

(MTA), que estos realizan la función de transferencia de datos de un ordenador a otro, de manera eficiente. Ellos pueden funcionar de varias formas: como servidores de otros servidores, recibe mensajes de otro servidor; como cliente de otros servidores, envían los mensajes o como intermediarios entre un *Mail Submision Agent* **(MSA)** y otro MTA.

Postfix[14]

Este es un servidor de correo de software libre, creado con la intención de que sea una alternativa más rápida, fácil de administrar y segura al ampliamente utilizado Sendmail, con el cual es totalmente compatible.

Entre sus ventajas se puede mencionar que posee soporte para distintas bases de datos, LDAP, mbox, maildir y dominios virtuales, además de manejar altos volúmenes de correo. Está compuesto de varios procesos que se comunican entre sí.

Es de fácil configuración mediante los dos ficheros main.cf y master.cf, permitiendo un uso sencillo de listas negras. También posibilita una fácil integración con herramientas externas, tales como: Postfixadmin, Mysql, Spamassassin y ClamAv (14).

Sendmail[15]

Este conocido enrutador de correos (MTA) es liberado bajo la licencia GNU *General Public Licens*e (GPL), es multiplataforma y su última revisión en prueba es la 8.14.5.Alpha0 liberada en el año 2010. Es un

[14] http://www.postfix.org

[15] http:// www.sendmail.com

programa increíblemente potente y altamente configurable. Para dichas configuraciones hace uso del fichero sendmail.cf.

Se señala su alto número de alertas de seguridad, además de no ser sencillo de configurar. Sus desarrolladores han querido hacer de este un producto muy flexible lo cual ha terminado por convertirlo en muy complejo. Aun cuando es uno de los MTA más utilizados en el mundo muy pocos administradores llegan a dominarlo en su totalidad a causa de que muchas de las características que en otros sistemas se mantiene a nivel de compilación en este son configurables (15).

Con el propósito de elaborar una propuesta que permita la configuración de un filtro de correo electrónico de forma sencilla, se propone la utilización del MTA Postfix para el servidor de correo. Debido a que posee una estructura flexible que permite la incorporación de los programas necesarios para realizar el filtrado de correo electrónico. Además, brinda las condiciones adecuadas para la integración con antivirus como el Clamav, actualmente utilizado por Smart Keeper, y la integración con otras herramientas muy útiles para la administración del proceso de filtrado.

1.3.3 Redirección de correos electrónicos

El Agente de entrega de correo es un software que acepta correo entrante de un Agente de Transporte y los distribuye a los buzones de los destinatarios. Para realizar el filtrado de correo estos programas son útiles en el proceso de redirección de los mensajes desde el agente de transporte, Postfix, hacia el sistema de filtrado. Por tanto se hace necesario definir una herramienta de este tipo. Entre los más conocidos actualmente se pueden encontrar Maildrop y Procmail.

Procmail[16]

Es un programa que permite procesar correos, de una forma sencilla. Separa los correos según determinados filtros, elimina correos *spam*, re-envía a otras cuentas y ejecuta programas. Procmail es configurable por cada usuario de manera personalizada editando el fichero .procmailrc ubicado en la carpeta *home* de cada usuario, este fichero se edita mediante el uso de expresiones regulares con el objetivo de filtrar o redireccionar correo a carpetas específicas.

Maildrop

Es un MDA con capacidad de filtrado programado en C++, suite de Courier. Sustituye al software de reparto de correo local. Este trabaja con buzones tipo mbox y maildir. A diferencia de Procmail, utiliza un lenguaje estructurado. Maildrop no guarda un mensaje de 10 MB en memoria lo salva a un fichero temporal y lo filtra directamente (16). Para establecer las reglas de filtrado es necesario configurar el archivo /etc/maildroprc el cual es de uso general para todos los usuarios de correo electrónico.

El sistema Procmail brinda amplias posibilidades para configurar sus funcionalidades. Además, la posibilidad de definir expresiones regulares en su mismo fichero de configuración, y la ejecucion de programas son características que se tuvieron en cuenta para decidir la integración de este *script* con el sistema de filtrado para lograr la redirección de los mensajes.

1.3.3 Funcionamiento del correo electrónico

Dirección de correo

La dirección de correo es un conjunto de palabras que identifican a una persona que envía y recibe mensajes, por tanto va a ser única y solo va a pertenecer a una sola persona.

---

[16] http://www.procmail.org/

En 1971, Ray Tomlinson incorporó el uso de la arroba (@) en las direcciones, a la izquierda de la arroba se encuentra el usuario de la persona a la que corresponde el correo y a la derecha el dominio al cual este pertenece. El usuario va a ser una combinación de letras, números y algunos signos según la elección del usuario. Ejemplo de una dirección: iemartinez@estudiantes.uci.cu, el usuario será iemartinez y el dominio al que pertenece es estudiantes.uci.cu.

Campos del mensaje

Destinatario (Para): dirección de correo al que va dirigido el mensaje. Estas pueden ser una o varias, en caso de ser varias puede ser una lista donde estén las direcciones separadas por coma.

También existen los campos:

Campo CC (*carbon copy*): todos los que estén en lista reciben este mensaje aunque este no va dirigido a ellos sino a quien está en el campo Para.

Campo CCO o **BCC** (*Copia de Carbón* Oculta o *Blind Carbon Copy*): una variante del **CC**, pero solo muestra el destinatario.

Asunto: una descripción breve que describe el tema a tratar en el cuerpo de correo. Si el mensaje es una respuesta el asunto suele empezar por Re: (abreviatura de responder o *reply* -en inglés-). Cuando el mensaje procede de un reenvío el asunto suele comenzar por RV: (abreviatura de reenviar) o Fwd: (del inglés *forward*).

Adjunto: archivos que se quieran adicionar al mensaje, que pueden ser cualquier tipo de archivo digital.

Cuerpo del mensaje: puede ser texto o incluir formato, sin límite de caracteres. Dentro también se puede incluir una firma personalizada del usuario.

Remitente (De: o From: -en inglés-): indica quién envía el mensaje. Puede aparecer el nombre de la persona o entidad que lo envía.

Si el mensaje es una respuesta el asunto suele empezar por RE: o Re: (abreviatura de responder o *reply* -en inglés-, seguida de dos puntos), estará definido en correspondencia con el idioma en que este escrito el mensaje.

Cuando el mensaje procede de un reenvío el asunto suele comenzar por RV: (abreviatura de reenviar) o Fwd: (del inglés *forward*), aunque a veces empieza por Rm: (abreviatura de remitir).

El envío de un mensaje puede realizarse de dos formas:

1. El usuario utiliza un programa de tipo Agente de Usuario de Correo, *Mail User Agent* (MUA) que se utiliza para crear y enviar, entre otras funciones, correos electrónicos. Este programa indica al servidor del usuario quién envía el correo y cuáles son los destinatarios, el mensaje es enviado a través del protocolo SMTP hacia el servidor.
2. El usuario usa un servidor de correo electrónico vía web, el cual se conoce como *Webmail* o correo web. Para acceder a él hay que autenticarse mediante una página web con un usuario y una contraseña.

Una vez y el mensaje se encuentra en el servidor el servicio MTA local al usuario inicial recupera este archivo e inicia la negociación con el servidor del destinatario para el envío del mismo, este consulta con su servidor quien es el encargado de gestionarlo, es decir, pregunta por el registro MX y lo envía al servidor correspondiente a través del protocolo SMTP. Estando ya el correo en el servidor del destinatario el correo puede verse de las dos formas antes descritas (17).

1.3.4 Formato de mensaje

El formato de un mensaje está regido según el Request For Comments 822[17], las partes de un mensaje son el encabezado, el cuerpo del mensaje y los adjuntos**,** cada campo de la cabecera consiste en una sola línea de texto ASCII que contiene el nombre del campo, dos puntos (:) y, para la mayoría de los campos un valor, se presentan aquí los principales campos (18):

To: Direcciones de correo electrónico de los destinatarios primarios.

Cc: Direcciones de correo electrónico de los destinatarios secundarios. En términos de entrega no existe diferencia con los destinatarios primarios.

Bcc: Direcciones de correo electrónico de las copias al carbón. Es como el campo anterior excepto que esta línea se borra de todas las copias enviadas a los destinatarios primarios y secundarios.

From: Persona o personas que crearon el mensaje.

Sender: Dirección de correo del remitente. Puede omitirse si es igual al campo anterior.

Received: Línea agregada por cada agente de transferencia en la ruta. La línea contiene la identidad del agente, la fecha y hora de recepción del mensaje, además de otra información que pueda servir para detectar fallos en el sistema de enrutamiento. Se añaden apiladas en la cabecera, a medida que se intercambia el correo.

Return-Path: Puede utilizarse para identificar una trayectoria de regreso al remitente.

Date: Fecha y hora de envío del mensaje.

Reply-To: Se utiliza cuando la persona que escribió el mensaje y la que lo envió no desean ver la respuesta.

---

[17] Sitio oficial donde se encuentra el RFC. http://www.ietf.org/rfc/rfc822.txt

Message-Id: Número único para referencia posterior a este mensaje. Suele estar compuesto por un número y la dirección de correo completa del usuario que lo envía.

In-Reply-To: Identificador del mensaje al que este corresponde.

References: Otros identificadores de mensaje.

Keywords: Claves seleccionadas por el usuario. Resumen corto del mensaje para exhibir en una línea.

Ejemplo de una cabecera:

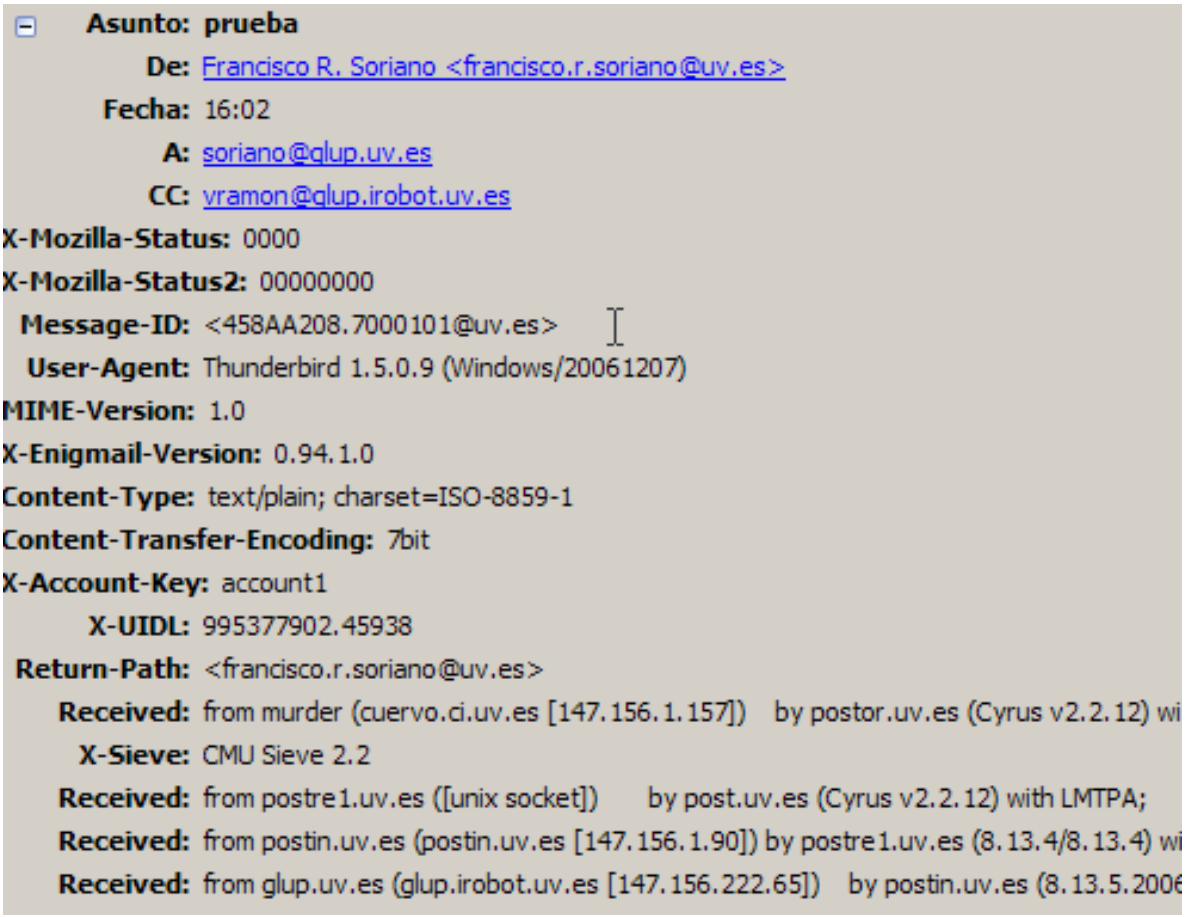


Imagen 1. Formato de un mensaje de correo electrónico.

El flujo de información generado por este servicio es incalculable si tenemos en cuenta que diariamente se envían cerca de 247 billones de mensajes (3) y de estos el 0.39 % son correos maliciosos, por lo que se puede decir que es una fuente inagotable de información y medio de socialización de los conocimientos. Pero este a la vez es un medio de propagación de virus, correos no deseados con publicidad falsa, o para obtener información bancaria, se estima que del tráfico de correo el 71 % son *spam* (2). Estas cifras son motivos suficientes para la implementación de filtros de contenido que controlen el flujo de la información y principalmente herramientas que permitan identificar los correos no deseados y los archivos infectados por virus.

1.4 Filtrado de correo electrónico

El filtrado de contenido de correo electrónico es un conjunto de técnicas mediante las cuales se analizan el tráfico SMTP haciéndolo pasar a través de diferentes filtros entre los que se encuentran: los filtros de contenido del correo, técnicas de clasificación de documentos e imágenes, filtros de *spam*, sistemas de listas grises u otro tipo de criterio que el desarrollador haya definido.

1.4.2 Técnicas de detección de correos *spam*

A partir del estudio realizado se han podido identificar un conjunto de técnicas anti *spam*. A continuación se describen las distintas técnicas para la detección de *spam*:

Listas grises (*greylisting*) es una técnica para el control de mensajes *spam*, mediante la cual el servidor de correo rechaza el mensaje enviado, considerándolo como posible *spam* y pide que sea reenviado. Esto sucede cuando no se tiene conocimiento alguno del remitente de correo, o sea, no se encuentran en las listas blancas o negras, o es considerado como sospechoso. Un servidor spammer normalmente envía mensajes en cantidades sin preocuparse si estos llegan o no, por tanto, no cumplirá con esta petición. Si

el mensaje es reenviado entonces se registrarán los datos siguientes: la dirección IP del remitente, dirección del remitente y dirección IP del destinatario.

Registro del Convenio de Remitentes (Sender Policy Framework) este método puede verificar si quien le está enviando el mensaje, está autorizado a enviar mensajes de ese dominio. El receptor (destinatario) consultará el DNS del dominio de origen en busca de este registro y verificará la procedencia antes de responder.

Aproximaciones colaborativas: en este tipo de técnica no se considera el contenido del mensaje, la detección de correo *spam* es llevada a cabo por un grupo de usuarios que comparten información sobre los mensajes *spam* (18).

Listas negras y blancas: se basan en reglas simples de exclusión de mensajes que han sido manipulados o vienen de dominios, redes o servidores de Internet. Con estas técnicas se puede clasificar e identificar gran cantidad de correo; pero también es muy fácil manipular este tipo de información. Existen aproximaciones similares a las reglas como son el análisis de los logs de servidores de correo electrónico, el análisis de los servidores de correo por los que ha viajado el mensaje o el análisis de la reputación de redes. A las listas negras se puede acceder mediante sitios web o archivos compartidos, estas se encuentran en forma de ficheros de texto que contienen remitentes o expresiones regulares acerca de direcciones de correo electrónico.

Por el contrario, las listas blancas contienen una numeración de equipos de los que nunca se debe desconfiar y poseen una confianza garantizada de una conexión anterior. Un gran inconveniente de esta técnica consiste en que los correos generados de forma automática, como los boletines de suscripción o las confirmaciones de suscripción nunca serán verificados (18).

Técnicas basadas en el análisis de contenido

Estas técnicas están comprendidas dentro de los métodos anti *spam* a la vez que dentro de ella existen otras formas de detección de correos *spam*. Estas tratan de determinar los atributos comunes en los mensajes a partir de una representación en forma de vector de las características de cada correo, para esto se selecciona una lista de palabras representativas de la legitimidad de los mensajes. Cada mensaje se representa con un vector de números reales o de valores lógicos que contiene, en cada posición, la frecuencia o presencia de los términos seleccionados.

Para la selección de los términos más representativos la técnica de selección de características más comúnmente utilizada se basa en el cálculo de la ganancia de información de cada término con respecto a los posibles valores del atributo a predecir (legítimo o *spam*). A continuación se resumen algunas de dichas técnicas (18):

- **Naïve Bayes:** es uno de los algoritmos más conocidos debido a su gran capacidad para representar de forma eficiente, distribuciones complejas de probabilidad. A pesar de que asume la independencia de los atributos de un correo electrónico lo cual en muchas ocasiones es poco realista, estudios realizados en la detección de correos *spam* han demostrado su gran efectividad.

  La probabilidad de que conocidos los valores que describen a un ejemplo *x*, éste pertenezcan a la clase v_j (donde v_j es el valor de la función de clasificación f(x) en el conjunto infinito V). Por el teorema de Bayes:

$$v_{MAP} = argmax_{v_j \in V} \frac{P(a_1, ...a_n|v_j)p(v_j)}{P(a_1, ...a_n)} = argmax_{v_j \in V} P(a_1, ...a_n|v_j)p(v_j)$$

Podemos estimar $P(v_j)$ contando las veces que aparece el ejemplo v_j en el conjunto de entrenamiento y dividiéndolo por el número total de ejemplos que forman este conjunto. Para estimar el término $P(a_1, ...a_n|v_j)$, es decir, las veces en que para cada categoría aparecen los valores del ejemplo *x*, se debe recorrer todo el conjunto de entrenamiento. Este cálculo resulta imposible para un conjunto grande de ejemplos por lo que se hace necesario simplificar la expresión. Para ello se presenta la hipótesis de independencia condicional con el objeto de poder factorizar la probabilidad.

- **Máquinas de Vectores de Soporte (SVM, Support Vector Machines):** su mecanismo de aprendizaje se basa en la idea de minimización de riego estructural (SRM, Structural Risk Minimization). Este algoritmo garantiza un buen nivel de generalización a partir de los datos de entrada. Han sido aplicados con mucho éxito en el campo de la detección de correo *spam*, por ser muy apropiado para la categorización de texto. Con el SVM no se necesita realizar selección de términos previa, debido a que su capacidad de aprendizaje no se degrada a medida que se incorporan nuevas características.

  La principal ventaja consiste en la gran velocidad de su tiempo de ejecución, aunque por otro lado tiene en su contra el hecho de que consumen mucho tiempo de entrenamiento si el fichero contiene muchos correos.

- **Sistemas de razonamiento basado en casos (CBR, Case-Based Reasoning):** estos sistemas son capaces de actualizar la base de casos cada vez que llega un nuevo mensaje. Al contrario de otros modelos no necesita reconstruir el modelo para incorporar un nuevo conocimiento, la base de casos se actualiza continuamente y el nuevo problema planteado se hace disponible para posteriores procesos de inducción dando la posibilidad

de realizar una re-selección de las características que puedan ser más idóneas para detectar los mensajes *spam*.

Una de sus principales ventajas consiste en que trabaja con grandes cantidades de datos de entrenamiento. Al integrar grandes cantidades de datos continuamente los sistemas CBR ofrecen técnicas de gestión de la base de casos para eliminar el ruido y los datos redundantes y gestionar así eficientemente el tamaño de la base de casos.

A partir del estudio de las técnicas antes descritas se ha podido determinar que el uso de métodos como el de listas negras, blancas y aproximaciones colaborativas será muy beneficioso para el desarrollo del filtro de correo electrónico. Por tanto, en la propuesta de filtrado de correo electrónico, se decide utilizar el el sistema SpamAssassin e implementar varios de los métodos antes mencionados.

1.5 Conclusiones

En este capítulo se han abordado elementos teóricos para sustentar la elaboración de la solución al problema inicialmente planteado. Debido a que los programas de filtrado de correo electrónico analizados son privativos y no distribuyen el código fuente de sus soluciones, surgió la necesidad de elaborar una propuesta de filtrado de correo electrónico para Smart Keeper, la cual utilice algunas de las mejores características de los sistemas estudiados.

Se analizaron los conceptos fundamentales del funcionamiento del correo electrónico y temas relacionados al filtrado. Se seleccionó como servidor de correo electrónico Postfix. Con el objetivo de redireccionar los mensajes desde el servidor hacia el sistema de filtrado se eligió el programa Procmail. Se ha seleccionado para realizar el filtrado anti spam el programa SpamAssassin, además de hacer uso de listas blancas y negras para fortalecer el filtrado.

Capítulo 2. Solución propuesta y validación

Introducción

Smart Keeper es un filtro de contenido web que permite aplicar un conjunto de políticas de navegación en una institución, con el fin de regular el acceso de los usuarios a Internet. Está basado en el uso e integración del servidor proxy cache Squid, el servidor web Apache2, el servidor de base de datos PostgreSQL, entre otros componentes libres. En este capítulo se hará una descripción de los elementos a incluir en la propuesta de filtrado de correo electrónico así como la validación de la misma.

2.1 Motor de Clasificación Inteligente por Contenido (MOCIC)

Se ha decidido que en un futuro se integre MOCIC a Smart Keeper con el objetivo de que este último posea una herramienta de categorización automático de contenidos. El sistema MOCIC tiene como propósito general recibir información digitalizada para su clasificación en categorías prestablecidas (19). Este se basa, principalmente, en la clasificación de información contenida en la Web; aunque está concebido para apoyar tareas como el Filtrado de Contenidos, la Organización de Documentos, el Seguimiento de Noticias, el Procesamiento Inteligente de Encuestas, el Seguimiento de objetos en videos, entre otras. Este sistema está compuesto por varios módulos:

- Clasificador de Texto
- Clasificador de Rostros
- Clasificador de Desnudez
- Clasificador de Objetos
- Clasificador de Enlaces

- Decisor
- Depósito
- Controlador
- Suministrador de Depósito

Filtrar el correo electrónico no cataloga como uno de los objetivos del proyecto MOCIC. Sin embargo, las funciones que este realiza se consideran un caso particular dentro del filtrado de contenidos y, por tanto, algunos de los módulos desarrollados por dicho proyecto son de gran utilidad a la hora realizar análisis del contenido de un correo electrónico. A continuación se hará una breve descripción de los módulos que se proponen utilizar para la categorización de los contenidos existentes en un mensaje de correo electrónico:

Módulo Clasificador de Texto: El objetivo fundamental de este módulo es determinar el idioma en el que ha sido redactado y las categorías de contenido previamente definidas asociadas a un texto. En el caso específico del correo electrónico dicho texto puede ser el cuerpo del mensaje o un documento adjunto.

Módulo Clasificador de Desnudez: Este módulo se encarga de determinar la existencia o no de desnudez en las imágenes. Estas pueden estar embebidas en el cuerpo del correo electrónico o como un archivo adjunto.

Módulo Clasificador de Enlaces: Este módulo tiene como función determinar las categorías asociadas a los enlaces detectados en un documento. Para ello se analizará el cuerpo del correo en busca de enlaces.

Módulo Decisor: Este módulo tiene como función recibir por parte de su homólogo Controlador toda la información proveniente de los módulos clasificadores, y devolverle la categoría más probable a la que pertenece el documento. Para la utilización de este módulo en el sistema de filtrado de correo electrónico será necesario hacer modificaciones en su actual comportamiento como es redefinir el lugar de ubicación de los contenidos a categorizar.

2.3 Funcionalidades de Smart Keeper en función del filtrado de correo electrónico

Smart Keeper es un sistema de filtrado de contenido web, sin embargo algunos de sus funcionalidades pueden ser aprovechadas en el filtrado de correo electrónico. Para un mejor entendimiento se partirá de explicar el filtrado del sistema Smart Keeper mediante las políticas de navegación que define. Dichas políticas son aplicadas a usuarios o grupos de usuarios. Están constituidas por categorías que agrupan URLs. Además aplican reglas para la restricción de ficheros, el chequeo de virus y otras reglas que regulan la navegación. Este principio de funcionamiento se puede extender al filtrado de correo electrónico, de forma que el sistema permita o deniegue el envío o recepción de mensajes en dependencia de políticas definidas para este fin. Las reglas mencionadas se definen mediante módulos que se integran a la interfaz de administración del sistema.

A continuación se describen los módulos del sistema que recogen dichas reglas y que se utilizarían para el filtrado del correo electrónico.

Restricciones de ficheros: Permiten regular el acceso a determinados ficheros según su MIME, extensión y tamaño. Esta regla se usaría para analizar los archivos adjuntos de un correo electrónico.

Chequeo de virus: La comprobación de virus permiten evitar la descarga de ficheros infectados según su MIME, extensión y tamaño. Los adjuntos de los correos serían revisados y en dependencia del análisis el administrador podrá decidir si solo elimina el adjunto o el mensaje completo.

Categorías: Las URLs almacenadas en el sistema constituyen la base del proceso de filtrado. Cada URL puede tener asignada una o varias categorías. A cada mensaje se le asociará una categoría que dependerá del resultado obtenido de la categorización por parte de los módulos de MOCIC Y MOCICE del texto del cuerpo del mensaje, de los archivos adjuntos, de las imágenes y de los enlaces contenidos en el mensaje.

2.4 MOCICE

Este módulo permite agrupar a los usuarios por tópicos, es decir, por los temas que tratan en sus correos. También permite determinar las comunidades a las que ellos pertenecen, las cuales están constituidas por el conjunto de usuarios que han establecido comunicación entre ellos mediante correos. En ambos casos el sistema actualmente no asigna una categoría, solo agrupa los mensajes. La detección no se hace actualmente en tiempo real. Además, se dispone de un sitio web a través del cual se puede tener acceso a los mensajes analizados, datos del usuario remitente, la cantidad de comunidades a las que este pertenece y otros datos de interés para el administrador del sistema.

2.5 Arquitectura de la solución

Partiendo de los elementos definidos en el Capítulo 1, se utilizará el MTA Postfix para hacer uso de su flexibilidad, y a la vez, fácil configuración. Con motivo de redireccionar los correos del servidor Postfix hacia el sistema de análisis del mensaje o sistema de filtrado, se hará uso del programa Procmail.

Una vez que el mensaje se encuentra en el sistema de filtrado este lo enviará hacia el programa SpamAssassin para realizar el análisis anti *spam.* Luego se procederá a realizar el análisis de cada una de sus partes, dicho proceso se puede dividir en 3 fases:

- **Fase 1:** Análisis de dirección de correo electrónico.
- **Fase 2:** Análisis de adjuntos del correo electrónico.
- **Fase 3:** Análisis del cuerpo del mensaje del correo electrónico.

En el análisis de direcciones el administrador del sistema podrá determinar si enviar los mensajes categorizados como *spam* hacia la carpeta de *spam* del usuario o bloquear el mensaje. En caso de no

serlo se verificará su presencia en las listas blancas o negras, y se analizará por el módulo de Expresiones Regulares del sistema Smart Keeper.

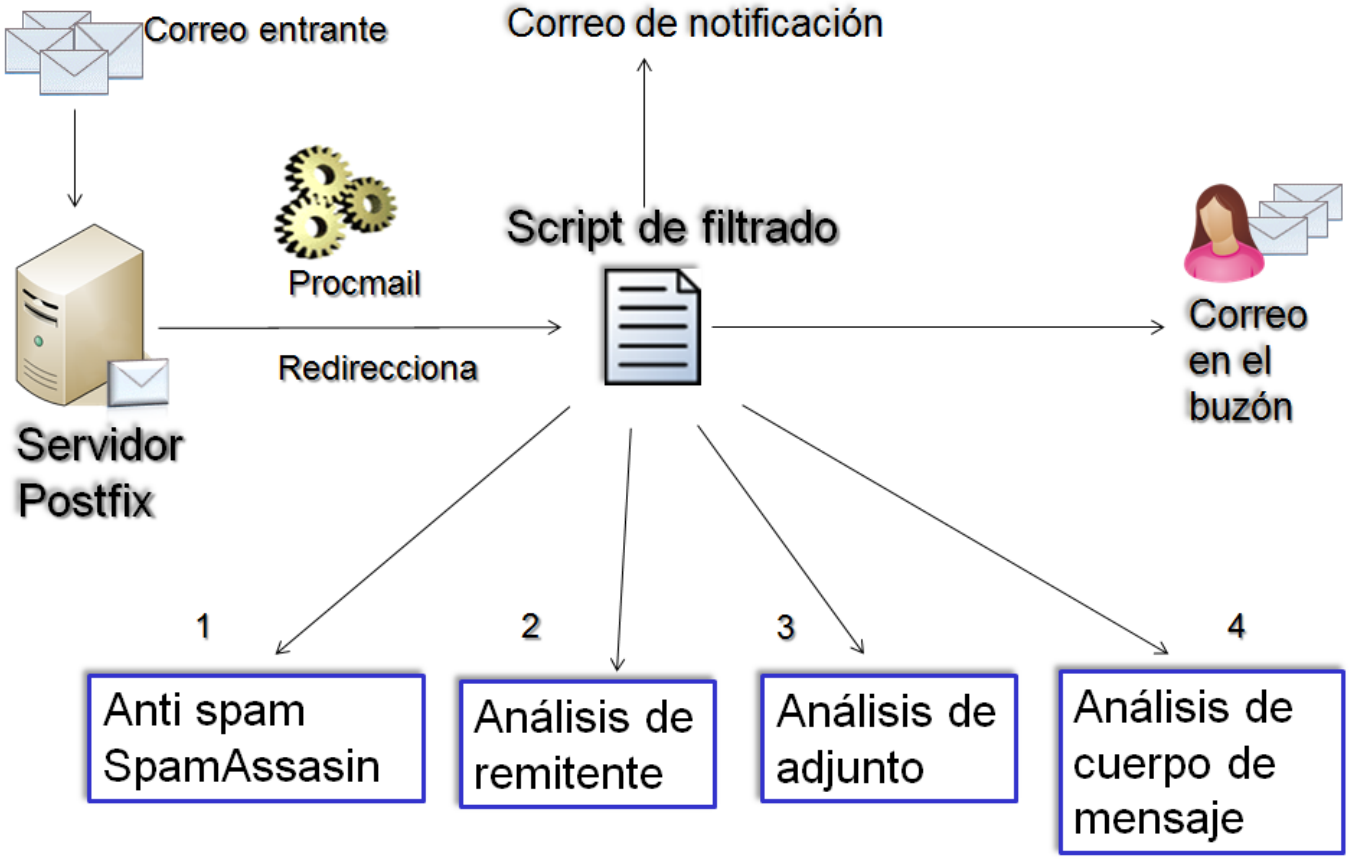
El análisis de adjuntos se llevará a cabo por los módulos de Smart Keeper Extensiones de Ficheros y Chequeo de Virus, seguidamente se analizarán los adjuntos por los módulos de MOCIC Análisis de Texto y Análisis de Desnudez. Dicho análisis se realizará siguiendo estrictamente el orden con que fueron mencionados los módulos.

En caso de que alguno de los adjuntos resulte bloqueado por el módulo Extensiones de Ficheros o se detecte que está infectado por el módulo Chequeo de Virus no se proseguirá con el análisis, se bloqueará todo el mensaje o solamente el adjunto en dependencia de la configuración realizada por el administrador.

Si se procede con el análisis de los módulos de MOCIC, se determinará si el mensaje será bloqueado en dependencia de las políticas definidas por el sistema Smart Keeper para la dirección de correo del usuario.

El módulo de MOCICE permitirá generar estadísticas sobre las cantidades de usuarios relacionados a un tema o los grupos de usuarios relacionados entre sí. Este no formará parte del proceso de filtrado pero ampliará el alcance de dicho producto relacionado al procesamiento de los mensajes de correo electrónico.

Para la administración de los análisis antes descritos será necesario desarrollar un módulo que se integre a la interfaz web de Smart Keeper, a fin de aprovechar esta herramienta de dicho sistema. El sistema ha de permitir al administrador configurar el proceso de filtrado de forma que permita la utilización de cada uno de los sub-módulos según sus necesidades. A continuación se muestra un esquema del proceso de filtrado de correo electrónico. Véase Figura 2 en Anexos una imagen más detallada.



A continuación se describirá con más detalles las 3 fases del análisis antes mencionado.

2.5.1 Fase 1: Análisis de dirección de correo

El análisis de dirección de correo se le realizará a la dirección del remitente que podrá estar sujeta a alguna política definida en Smart Keeper. Este sistema debe permitir la definición de listas de denegación

(listas blancas o listas negras) haciendo uso de la información almacenada en la base de datos, específicamente la dirección de correo electrónico de un usuario.

Las políticas dirigidas a estos usuarios pueden ser definidas por el tipo o tamaño de los archivos adjuntos o las categorías de los contenidos a tratar en el mensaje según un determinado horario. Haciendo uso del módulo de expresiones regulares se podrán crear reglas y modificar las listas de los servidores de correo permitidos o bloqueados.

El filtro de correo electrónico también podrá hacer uso de proveedores gratis de listas RBL como son los sitios www.njabl.org y www.spamcop.net a los cuales se pueden tener acceso solo con una subscripción al sitio.

2.5.2 Fase 2: Análisis de adjuntos

El análisis de adjuntos, se hará a través del módulo de Restricción de ficheros de Smart Keeper en el cual se podrá bloquear o permitir los tipos de ficheros en la sección de Administración MIME y extensiones donde se le podrá aplicar ciertas restricciones a un conjunto de usuarios, a una política (por ende a los usuarios pertenecientes a dicha política) o a todos en general. En este módulo se pueden bloquear o permitir las extensiones en un rango de tiempo o según el tamaño de los ficheros. Las políticas relacionadas con los adjuntos también podrán estar definidas según su categoría que bien pueden estar predefinidas en el sistema Smart Keeper o ser el resultado del análisis realizado por los módulos de MOCIC.

Mediante el módulo de Chequeo de antivirus, que hace uso del antivirus Clamav, se podrán chequear ficheros, escanear según su extensión MIME, y se configurará el tamaño de los ficheros que se van a escanear.

2.5.3 Fase 3: Análisis de cuerpo de correo

El análisis del cuerpo de correo se llevará a cabo a través del módulo de Categorización de Texto de MOCIC. En dicho módulo se analiza el mensaje detectándose las comunidades a las que pertenece, es decir, quedan agrupados según los usuarios con los que este se relaciona o el tema común entre ellos. Para estos usuarios también pueden estar definidas un conjunto de políticas ya sea a un usuario en particular o de modo global para un conjunto de usuarios.

La integración de Smart Keeper con los módulos de MOCIC de Clasificador de texto, imágenes y enlaces permitirá asignarles una categoría a los mensajes, la cual será útil para la aplicación de políticas a los usuarios. Para visualizar los resultados de estas clasificaciones deberá existir un sitio web donde se pueda tener acceso a las categorías existentes y las estadísticas referentes a ellas y a los mensajes clasificados.

El módulo MOCICE permitirá agrupar los mensajes por comunidades de usuarios y por los temas tratados en ellos. Permitiendo generar estadísticas que puedan facilitar el filtrado de correo electrónico a la hora de establecer las políticas de uso de correo electrónico. Este análisis no se hará en tiempo real.

2.6 Validación

Es necesario evaluar la propuesta descrita en los epígrafes anteriores para verificar que esta contiene los elementos necesarios y cumple los objetivos trazados en dicho trabajo. En este epígrafe se hace un análisis de los métodos que permiten hacer este tipo de evaluaciones, con vistas a la selección y aplicación del más adecuado.

2.6.1 Métodos de Evaluación

Método de consulta a expertos (Delphi): Este trata de lograr un consenso de opiniones expresadas individualmente por un grupo de expertos en el tema, por medio de la iteración sucesiva de un

cuestionario retroalimentado de los resultados promedio de la ronda anterior, aplicando cálculos estadísticos (20).

Método de validación práctica: Consiste en la obtención, comparación y análisis de resultados obtenidos en el transcurso de la investigación, tras aplicar de manera práctica el procedimiento en varios proyectos (21).

Método de grupo focal: Este método se basa en la selección de un conjunto de personas con abundantes conocimientos del tema. Deben ser expertos o especialistas, de distintos niveles y categorías. Estos se reúnen en un lugar a una hora determinada donde se discute, en forma de grupo-debate dirigido por los autores, lo que se quiere conocer sobre el procedimiento (22).

A partir de los métodos descritos anteriormente se selecciona el método Delphi. El método de validación práctica no es viable debido a que el proceso de filtrado de correo electrónico requeriría la implementación de módulos adicionales en el sistema Smart Keeper que aún no han sido desarrollados, lo cual no es posible debido al tiempo disponible y al alcance definido para este Trabajo de Diploma. El método de grupo focal se desestima debido al poco tiempo disponible para su instrumentación. Los expertos seleccionados poseen responsabilidades de trabajo que dificultan ajustar la fecha de la reunión de forma que todos estén presentes.

2.6.2 Evaluación de la viabilidad de la propuesta sobre la base del criterio de expertos (Método Delphi)

Esta técnica proporciona el conocimiento del grupo de expertos sobre el tema de investigación y es útil como herramienta exploratoria para el pronóstico tecnológico, permitiendo conocer la información de consenso en un grupo.

Consiste en la aplicación de un cuestionario a cada uno de los posibles expertos, con vistas a medir su coeficiente de competencia, obteniendo los que presentan aptitudes para dar respuesta a cada una de las interrogantes. Las conclusiones del análisis de las respuestas se traducen en un segundo cuestionario, que de nuevo se remite al grupo de expertos. El método suele dividirse en tres etapas o fases fundamentales:

Fase inicial: Es donde ocurre la formulación del problema. Las preguntas deben ser precisas, cuantificables e independientes.

Fase exploratoria: En esta fase ocurre la selección de los expertos con toda la serie de parámetros que la misma requiere.

Fase final: Es donde ocurre el desarrollo práctico y la explotación de los resultados.

El método Delphi se caracteriza por:

- **Anonimato:** No es necesario que los expertos se conozcan, lo cual garantiza que puedan emitir sus respuestas sin llegar a confrontaciones e impide que el criterio de un experto sea influenciado por la reputación de otro o por el peso de oponerse a la mayoría. Es posible que el gestor de la encuesta identifique a cada participante con sus respectivas respuestas.

- **Iterativo:** Pueden ser realizadas tantas iteraciones como sean necesarias para obtener datos concisos.

- **Retroalimentación controlada:** Luego de cada ronda de preguntas, las respuestas son tabuladas y procesadas de manera que en la siguiente ronda puedan ser evaluados los resultados de la anterior, así como las razones dadas por cada respuesta y su dispersión del promedio.

- **Respuesta estadística del grupo:** Entre rondas de preguntas, la información obtenida se procesa por medio de técnicas estadístico-matemáticas, las cuales dotan al investigador de un instrumento objetivo y concreto para apoyarse a la hora de tomar una decisión final.

2.6.3 Selección de los Expertos

Los expertos son claves en la aplicación el método Delphi, pues de sus opiniones depende la validez de la propuesta. Para seleccionar los expertos se consideraron las siguientes condiciones:

- Debe ser graduado de nivel superior.
- Debe poseer conocimiento acerca de los sistemas de filtrado de contenido.
- Debe estar vinculado a la administración servicios de correo electrónico.
- Debe poseer conocimiento acerca de los sistemas de filtrado de correo electrónico.
- Debe contar con al menos tres años de experiencia en el tema.

No necesariamente un mismo especialista tiene que dominar todos los temas.

Luego, a partir de estos criterios fueron seleccionados seis expertos:

- Cinco de ellos con experiencia media/alta en administración de servidores de correo.
- Todos poseen conocimientos medios/altos acerca de los sistemas de filtrado de contenido web.
- Tres de ellos poseen experiencia sobre los sistemas de filtrado de correo electrónico.

Con vistas a determinar la competencia de los expertos seleccionados, se elaboró un cuestionario en el cual cada experto valora sus conocimientos. Los resultados de esta encuesta se reflejan en la siguiente tabla. Dicha valoración se basa en una puntuación que va desde el 1 al 5.

Criterio/Evaluación	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6
Administración de servidores de correo electrónico	3	1	4	4	4	5
Sistemas de filtrado de contenido web	4	3	3	5	4	4
Sistemas de filtrado de correo electrónico	3	3	2	1	3	4

Todos los expertos son especialistas en uno de los criterios definidos y dominan al menos un segundo criterio, por lo cual se consideran aptos para llevar a cabo la validación de esta propuesta.

2.6.4 Conformación del cuestionario

Para la conformación del cuestionario se utilizaron los requisitos de calidad definidos para su evaluación de forma automática. A partir de estos fueron elaborados los aspectos en el cuestionario:

- Selección adecuada de las herramientas a utilizar en el proceso de filtrado.
- Selección adecuada de las herramientas a utilizar en el proceso de filtrado de correos *spam*.
- Selección adecuada de las herramientas para realizar la categorización de las partes de un mensaje.
- Descripción correcta del proceso.

- Idoneidad de la propuesta de acuerdo a las condiciones y los objetivos del filtrado de correo electrónico en el sistema Smart Keeper.

Una vez definidos estos aspectos, se elaboraron las preguntas que fueron incluidas en la encuesta y que tienen por objetivo evaluar tanto los programas como el proceso de filtrado propuesto (Véase Anexo 1).

2.6.5 Análisis estadístico de los datos

En esta fase se lleva a cabo el análisis estadístico de los datos que fueron obtenidos luego de aplicar la encuesta a los especialistas. Una vez aplicada la encuesta a los especialistas, los datos obtenidos son procesados utilizando el factor de Kendall para determinar el grado de concordancia entre sus opiniones.

Para obtener este valor fue utilizada la herramienta *Statistical Package for the Social Sciences*[18] (SPSS). Este programa estadístico informático es utilizado en las ciencias sociales y las empresas de investigación de mercado. Posibilita el trabajo con bases de datos de gran tamaño. Los resultados obtenidos por dicha herramienta se presentan a continuación:

Test Statistics

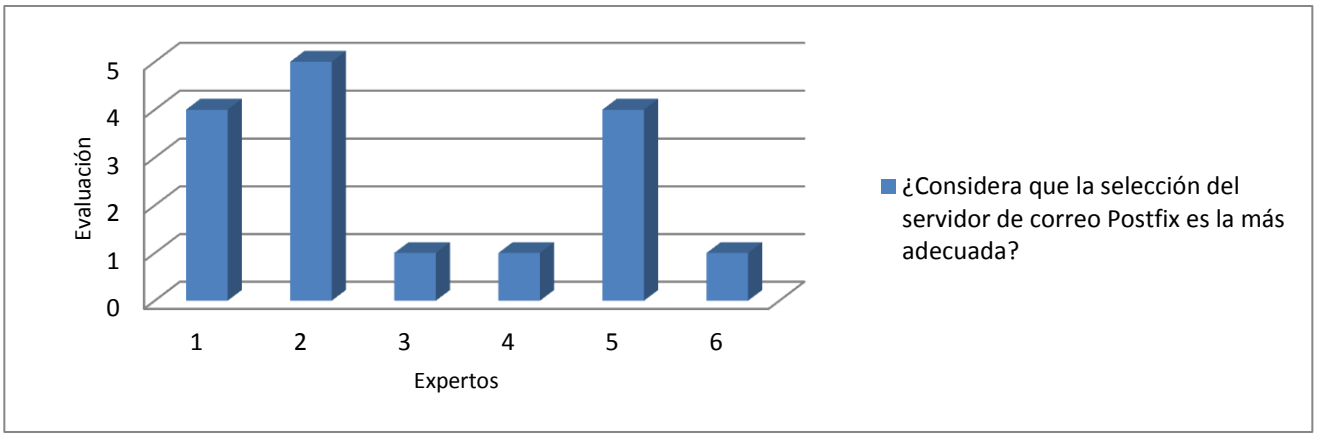
N	6
Kendall's W(a)	0,176
Chi-Square	6,343
df	6
Asymp. Sig.	0,386

a Kendall's Coefficient of Concordance

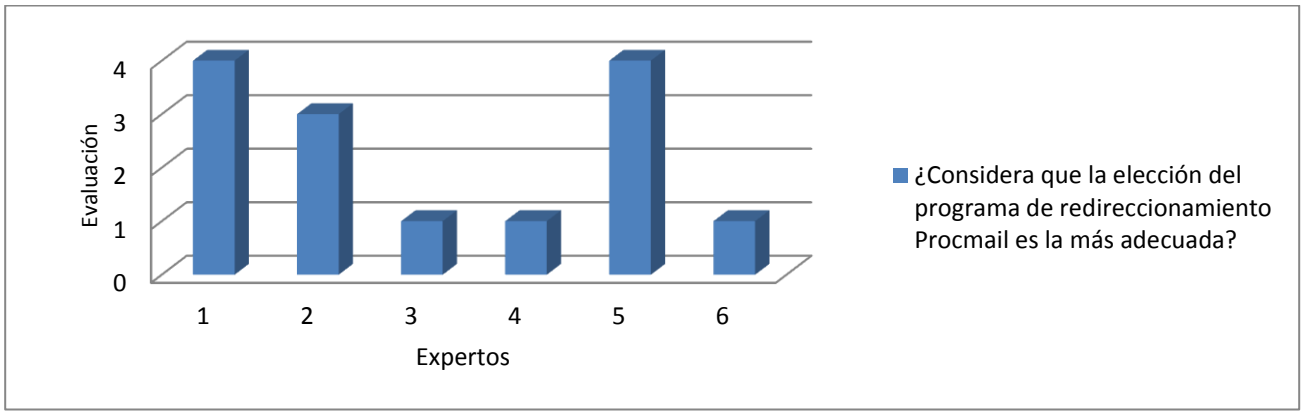
[18] Traducción de la autora: Paquete estadístico para las Ciencias Sociales.

Los valores del coeficiente de Kendall deben oscilar entre cero y uno (0 < k < 1); los valores cercanos a uno representan concordancia entre los expertos. Una vez calculado se obtiene que k = 0.18, lo cual representa que los expertos coinciden en un 18 %.

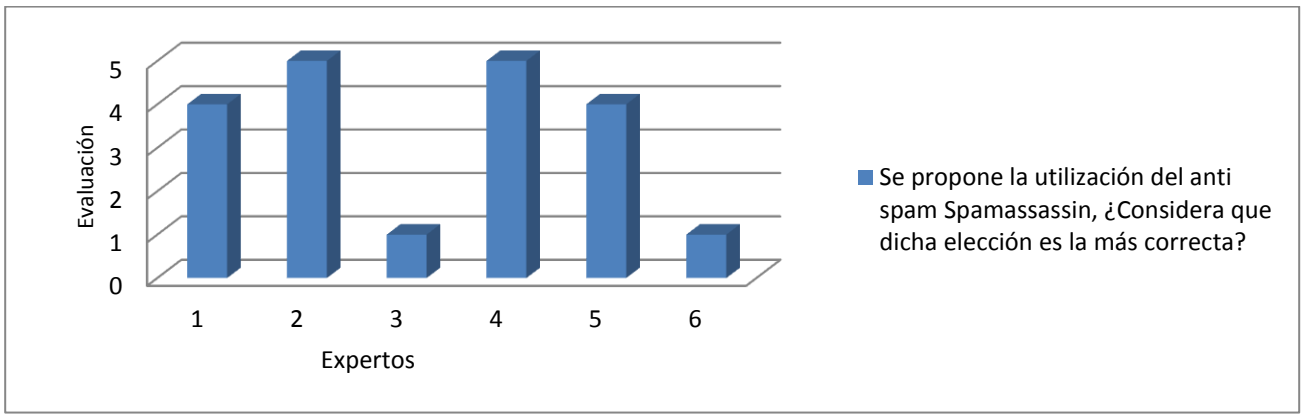
Para comprobar la validez de la propuesta se busca el Chi-cuadrado, tabulado en la tabla del percentil de la distribución Chi-cuadrado con un nivel de significación yn-1 grados de libertad, representado por $\chi^2_{tab} = \chi^2_{\alpha;n-1}$ y se compara con el Chi-cuadrado calculado por la herramienta, si $\chi^2_{cal} > \chi^2_{tab}$ entonces la propuesta es válida. Comparando Chi-cuadrado calculado = 6.343 con Chi-cuadrado tabulado =1.145 se tiene que $\chi^2_{cal} > \chi^2_{tab}$ por lo tanto podemos afirmar que los expertos coinciden en que la propuesta es válida. A continuación se muestran gráficos que exponen los resultados de cada uno de los criterios establecidos.



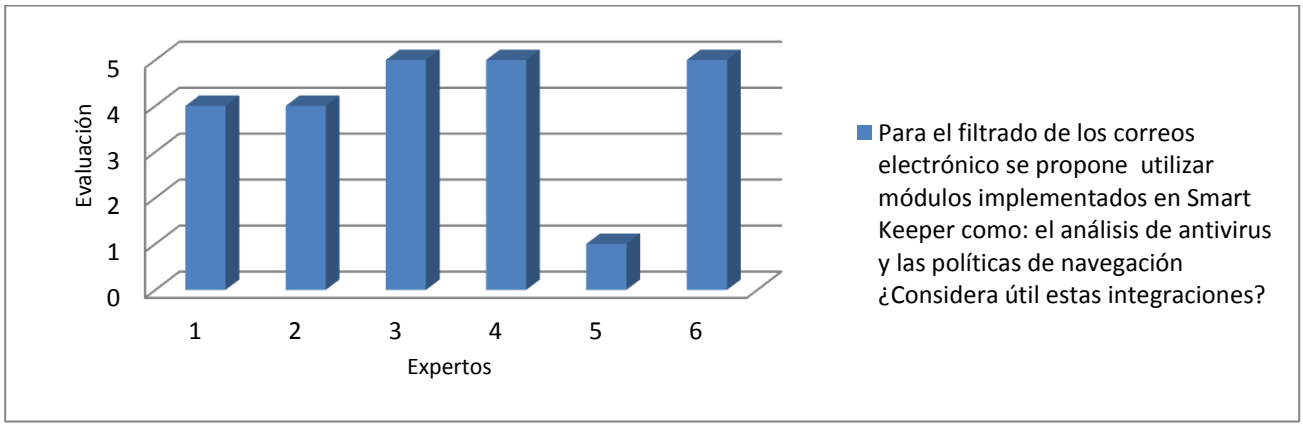
La gráfica muestra la valoración de los expertos respecto a la selección del servidor de correo Postfix, promediando para un 2,66.



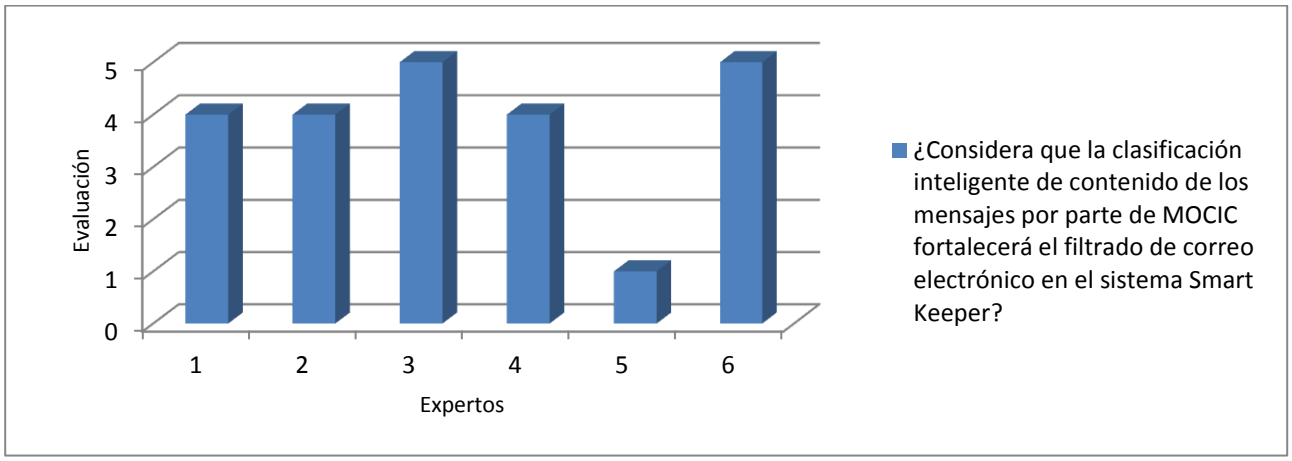
La gráfica muestra la valoración de los expertos acerca de la selección del programa de redireccionamiento Procmail, coincidiendo sus respuestas en un 2,33.



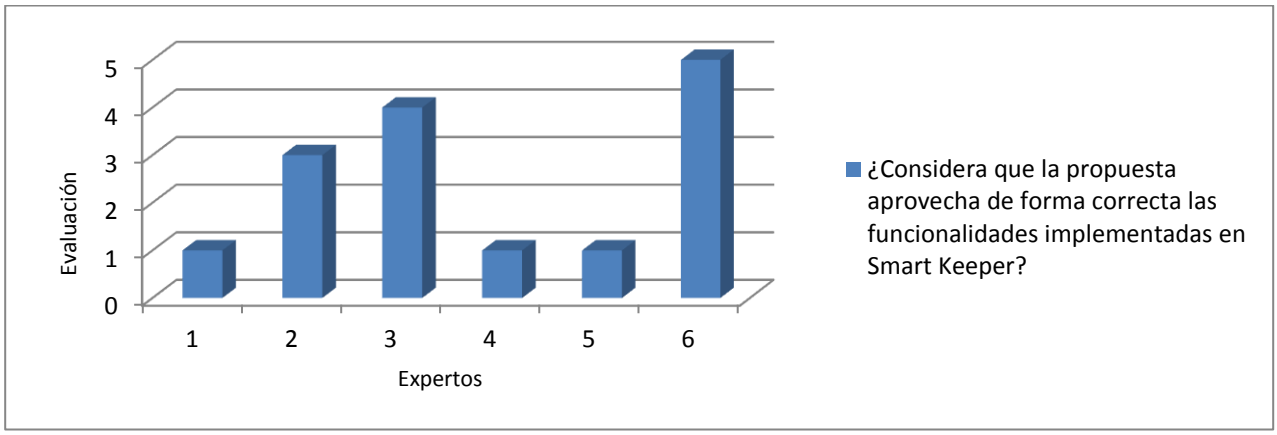
La gráfica muestra la valoración de los expertos respecto al uso del programa anti *spam* SpamAssassin, coincidiendo sus respuestas en un 3,33.



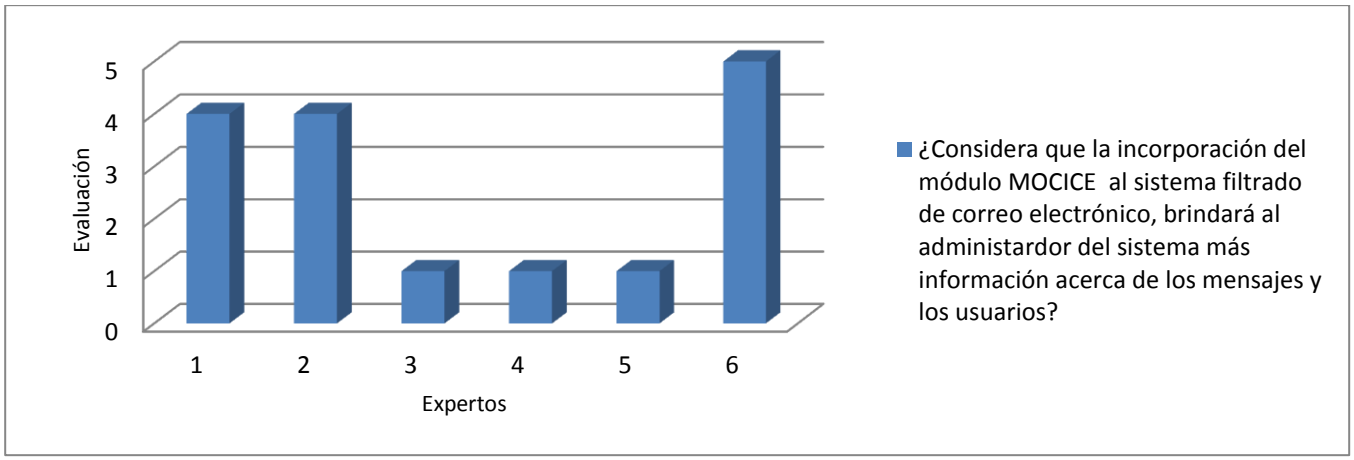
La gráfica anterior muestra la valoración de los expertos acerca de la integración de los módulos de Smart Keeper con el filtrado de correo electrónico, promediando 4.



La gráfica muestra la valoración de los expertos respecto a los beneficios que reportará la inclusión de MOCIC en el sistema Smart Keeper, promediando un 3,83.



En la gráfica se muestra la valoración de los expertos respecto al aprovechamiento de las funcionalidades implementadas actualmente en Smart Keeper con el objetivo de filtrar los correos electrónicos, promediando un 2,5.



La gráfica muestra la valoración de los expertos respecto a las ventajas brindadas por MOCICE para el sistema de filtrado de correo electrónico en el sistema Smart Keeper, promediando un 2,66.

En la mayoría de los casos el promedio de aceptación de la propuesta estuvo por encima de 2,5. En el caso de las preguntas que no fueron contestadas se le asignó el valor 1, debido a que los expertos no eran especialistas en todos los temas. No obstante cada uno de ellos estuvo de acuerdo en con que se habían analizado cada uno de los elementos necesarios para realizar el filtrado de correo electrónico. Por tanto la propuesta queda validada por el método de expertos (Véase Anexo 3).

2.7 Conclusiones

Al término del presente capítulo se evidenció la necesidad de incorporar al sistema Smart Keeper un filtro de correos *spam*, así como otras características que puedan ser requeridas para instrumentar la integración con el sistema de filtrado de correo electrónico. Además se evidenció la necesidad de una re-adaptación de algunos de los módulos de MOCIC y MOCICE, en función de los datos que tomará de cada mensaje y del formato en que puedan aparecer. Una vez validada la propuesta descrita, se infiere la viabilidad de esta y la importancia de su incorporación al sistema Smart Keeper.

Conclusiones

Luego de analizar la posibilidad de utilizar una solución existente para dar respuesta a la problemática, se decidió elaborar una propuesta de filtrado de correo electrónico que fuese específica para el sistema Smart Keeper. Al finalizar la investigación se concluye que:

- El estudio del estudio del estado del arte permitió contar con los elementos teóricos necesarios para desarrollar la propuesta de filtrado de correo electrónico.
- El estudio de elementos teóricos a cerca del filtrado de correo electrónico permitió una mejor selección de las herramientas a incluir en la propuesta.
- La integración del software SpamAssassin al filtro de correo electrónico permitirá que dicho filtrado sea más preciso y eficaz.
- El uso de módulos de MOCIC permitirá que la solución propuesta cuente con una clasificación inteligente de contenidos, lo cual eleva en precisión los resultados.

Recomendaciones

- Lograr que el análisis del módulo MOCICE se realice en tiempo real.
- Lograr que el análisis del antivirus Clamav se realice en tiempo real.

Glosario de términos:

TCP/IP: *Transmission Control Protocol/Internet Protocol*, protocolo de control de la transmisión/protocolo de Internet.

MUA: *Mail User Agent*, Agente de usuario de correo.

SMTP: *Simple Mail Transfer Protocol*, protocolo simple de transferencia de correo electrónico.

DNS: *Domain Name System*, sistema de nombres de dominio.

Registro MX: *MX record*, registro de correo donde constan las direcciones que pertenecen a un dominio en concreto y la máquina que se corresponde con cada dirección. En este pueden aparecer varias máquinas estableciéndose prioridades entre ellas o balanceando la carga de correo por todas ellas.

POP3: *Post Office Protocol*, Protocolo de Oficina Postal versión 3.

IMAP*: Internet Message Access Protocol*, protocolo de Internet para el acceso a mensajes electrónicos.

Proxy: servidor de mediación, encargado, entre otras cosas, de centralizar el tráfico entre internet y una red privada, de forma que evita que cada una de las máquinas de la red interior tenga que disponer necesariamente de una conexión directa a la red. Al mismo tiempo contiene mecanismos de seguridad (cortafuegos) que impiden accesos no autorizados desde el exterior hacia la red privada.

IETF: *Internet Engineering Task Force*, en español Grupo Especial sobre Ingeniería de Internet, es una organización internacional abierta de normalización, que tiene como objetivos el contribuir a la ingeniería de Internet, actuando en diversas áreas, como transporte, encaminamiento, seguridad. La IETF es mundialmente conocida por ser la entidad que regula las propuestas y los estándares de Internet, conocidos como RFC.

Cookie: Es un archivo temporal que se almacena en nuestra computadora, y que conserva información de sitios web y la actividad en ellos, por ejemplo, el nombre de usuario. Estos archivos se guardan automáticamente.

HTTP: *Hypertext Transfer Protocol*, Protocolo de transmisión de hipertexto.

Wi-Fi: mecanismo de conexión de dispositivos electrónicos de forma inalámbrica.

ISP: Un proveedor de servicios de Internet, por la sigla en inglés de *Internet Service Provider* es una empresa que brinda conexión a Internet a sus clientes. Muchos ISPs también ofrecen servicios relacionados con Internet, como el correo electrónico, alojamiento web, registro de dominios, servidores de noticias, etc.

ASCII: acrónimo inglés de *American Standard Code for Information Interchange*, en español Código Estándar Estadounidense para el Intercambio de Información, es un código de caracteres basado en el alfabeto latino, tal como se usa en inglés moderno y en otras lenguas occidentales.

Proveedores de correo electrónico: empresa que ofrece este servicio de manera gratuita o pago, en la que asigna una dirección de correo única al cliente a la cual acceder mediante una contraseña.

MTA: *Mail Transport Agent*, también *Message Transport Agent*, Agente de Transporte de Mensajes.

MSA: *Mail Submision Agent*, es un programa informático o agente de software que recibe mensajes de correo electrónico desde un *Mail User Agent* y coopera con un *Mail Transport Agen*t para entregar el correo.

URL: Localizador de recursos uniforme, sigla en inglés de *Uniform Resource Locator*, es una secuencia de caracteres, de acuerdo a un formato modélico y estándar, que se usa para nombrar recursos en Internet para su localización o identificación.

Licencia Pública General de GNU: GNU *General Public License*, sus siglas del inglés GNU GPL, es una licencia creada por *la Free Software Foundation* en 1989 (la primera versión, escrita por Richard Stallman), y está orientada principalmente a proteger la libre distribución, modificación y uso de software.

RFC: *Request for Comments*, en español Petición de Comentarios, son una serie de notas sobre Internet, y sobre sistemas que se conectan a Internet, que comenzaron a publicarse en 1969. Cada una de ellas individualmente es un documento cuyo contenido es una propuesta oficial para un nuevo protocolo de la red que se explica con todo detalle para que en caso de ser aceptado pueda ser implementado sin ambigüedades.

Referencias bibliográficas

(1) **Internet Usage Statistics**. [En línea] 2011. [Citado el: 28 de noviembre de 2011.] http://www.internetworldstats.com/stats.htm.

(2) **Radicati**. [En línea] http://www.radicati.com/wp/wp-content/uploads/2010/04/Email-Statistics-Report-2010-2014-Executive-Summary2.pdf.

(3) **Pingdom**. [En línea] 2012. http://royal.pingdom.com/2012/01/17/internet-2011-in-numbers/

(4) **Top Ten Reviews**. [En línea] http://spam-filter-review.toptenreviews.com.

(5) **Spamfighter**. [En línea] http://www.spamfighter.com/Lang_ES/Product_SMTP.asp.

(6) **Top Ten Reviews.** [En línea] http://evaluacion-de-software-antivirus.toptenreviews.com.

(7) **Muñoz Mellid, Javier.** [En línea] http://stuff.gpul.org/2006_cripto/doc/2006_JCRIP_00_antivirus.pdf.

(8) **Diccionario de la Lengua Española**. [En línea] 2011. [Citado el: 14 de diciembre de 2011.] http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=internet.

(9) **Rodríguez Ávila, Abel.** Iniciación a la red Internet: Concepto, funcionamiento, servicios y aplicaciones de Internet.

(10) **Red Innova.** [En línea] 2011.

http://www.redinnova.com/wpcontent/uploads/2011/05/Encuestas_mayo_RedInnova.pdf.

(11) **Herrera Pérez, Enrique.** *Tecnologías y redes de transmisión de datos.*

(12) **Tecnología de Información y la Comunicación** – Correo Electrónico. [En línea] http://tutoriales.igluppiweb.com.ar/ecorreo.pdf.

(13) **Barceló Ordinas, José María, y otros, y otros.** Software libre. s.l. : Eureka Media, 2004.

(14) **Postfix**. [En línea] http://www.postfix.org/documentation.html.

(15) **Bravo Estrada, Diego.** [En línea] http://es.tldp.org/Tutoriales/doc-guia-sendmail/doc-guia-sendmail-html/.

(16) **Lo Huang, Jose Luis.** *Coloquio Email.* 2011.

(17) **Gálvez Rojas, Sergio y García Sucino, Ignacio.** Java a tope: JavaMail en ejemplos. [En línea] 2006. [Citado el: 20 de diciembre de 2011.] http://books.google.co.ve/books?id=sbEa668gGgUC&pg=PA3&dq=funcionamiento+de+correo+electronico&hl=es&sa=X&ei=LDbxTqeLLIbd0QHQ0MmXAg&ved=0CDMQ6AEwAA#v=onepage&q=funcionamiento%20de%20correo%20electronico&f=false.

(18) *Sistemas inteligentes para la detección y filtrado de correo spam: una revisión.* **Méndez, José R., y otros, y otros.** 34, Valencia, España : s.n., 2007, Vol. 11.

(19) **Gil, Yoandri Ril.** *Descripción de la Arquitectura MOCIC.* 2009.

(20) **Vélez Pareja, Ignacio.** *Decisiones empresariales bajo riesgo e incertidumbre.* s.l. : Editorial Norma, 2003.

(21) **Rafael.** Recopilación de información: métodos interactivos. [En línea] 2010 http://www.monografias.com/trabajos61/recopilacion-informacion-metodo-interactivo/recopilacion-informacion-metodo-interactivo.shtml.

(22) **Guinot, Cinta.** *Métodos, técnicas y documentos utilizados en Trabajo Social.* s.l. : Universidad de Deusto, 2009.

Anexos

Tabla 1

Regiones del mundo	Población (2011 Est.)	Usuarios 31 de Dic., 2000	Usuarios, Dic. 31, 2011	Crecimiento 2000-2011	% Uso mundial
África	1,037,524,058	4,514,400	**118,609,620**	2,527.4 %	5.7 %
Asia	3,879,740,877	114,304,000	**922,329,554**	706.9 %	44.0 %
Europa	816,426,346	105,096,093	**476,213,935**	353.1 %	22.7 %
Medio Oriente	216,258,843	3,284,800	**68,553,666**	1,987.0 %	3.3 %
Norte América	347,394,870	108,096,800	**272,066,000**	151.7 %	13.0 %
América Latina/ Caribe	597,283,165	18,068,919	**215,939,400**	1,037.4 %	10.3 %

Oceanía / Australia	35,426,995	7,620,480	**21,293,830**	179.4 %	1.0 %
TOTAL MUNDIAL	**6,930,055,154**	**360,985,492**	**2,095,006,005**	**480.4 %**	**100.0 %**

NOTAS: (1) Las Estadísticas de Usuarios Mundiales del Internet fueron actualizadas a Marzo 31, 2011. (3) Los datos de población se basan en cifras para 2009 del US Census Bureau. (4) Los datos de usuarios provienen de información publicada por Nielsen Online, ITU y de Internet World Stats. (6) Estas estadísticas son propiedad intelectual de Miniwatts Marketing Group, se pueden citar, siempre manifestando el debido credito y estableciendo un enlace activo a www.exitoexportador.com.

1. Cuestionario para la validación de la propuesta

En la siguiente tabla responda las preguntas, asignándole un criterio a cada una según convenga. **Nota:** C1 es mayor que C5.

No.	**Preguntas**	**Criterio del Experto**				
		C1	**C2**	**C3**	**C4**	**C5**
1	¿Considera que la selección del servidor de correo Postfix es la más adecuada?					
2	¿Considera que la elección del programa de redireccionamiento Procmail es la más					

	adecuada?					
3	Se propone la utilización del antie *spam* Spamassassin, ¿Considera que dicha elección es la más correcta?					
4	Para el filtrado de los correos electrónico se propone utilizar módulos implementados en Smart Keeper como: el análisis de antivirus y las políticas de navegación ¿Considera útil estas integraciones?					
5	¿Considera que la clasificación inteligente de contenido de los mensajes por parte de MOCIC fortalecerá el filtrado de correo electrónico en el sistema Smart Keeper?					
6	¿Considera que la propuesta aprovecha de forma correcta las funcionalidades implementadas en Smart Keeper?					
7	¿Considera que la incorporación del módulo MOCICE al sistema filtrado de correo electrónico, brindará al administrador del sistema más información acerca de los mensajes y los usuarios?					

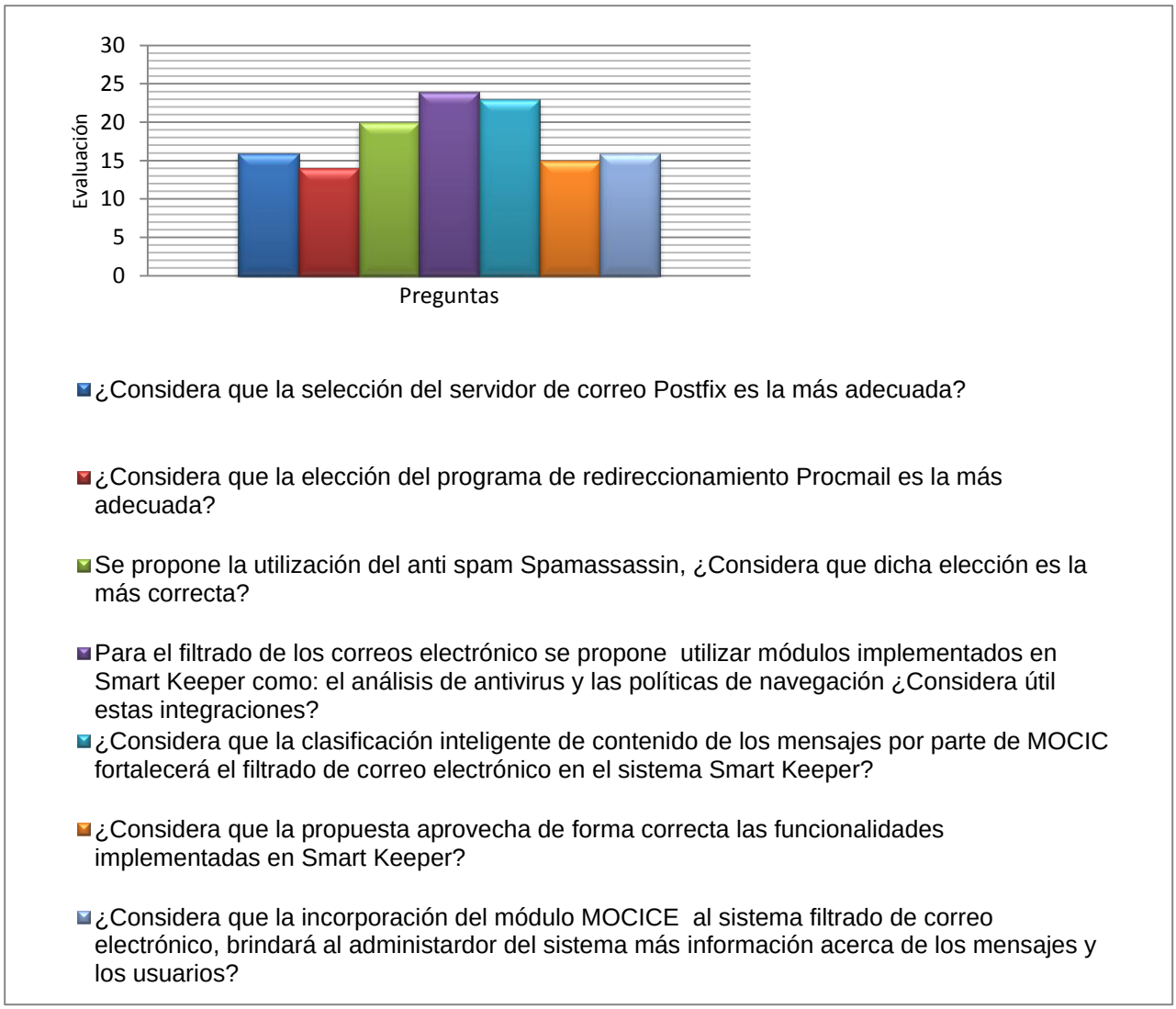
¿Qué otros elementos cree que se debieron tener en cuenta en la propuesta?

2. Encuesta realizada a los expertos para determinar sus conocimientos referentes a la propuesta

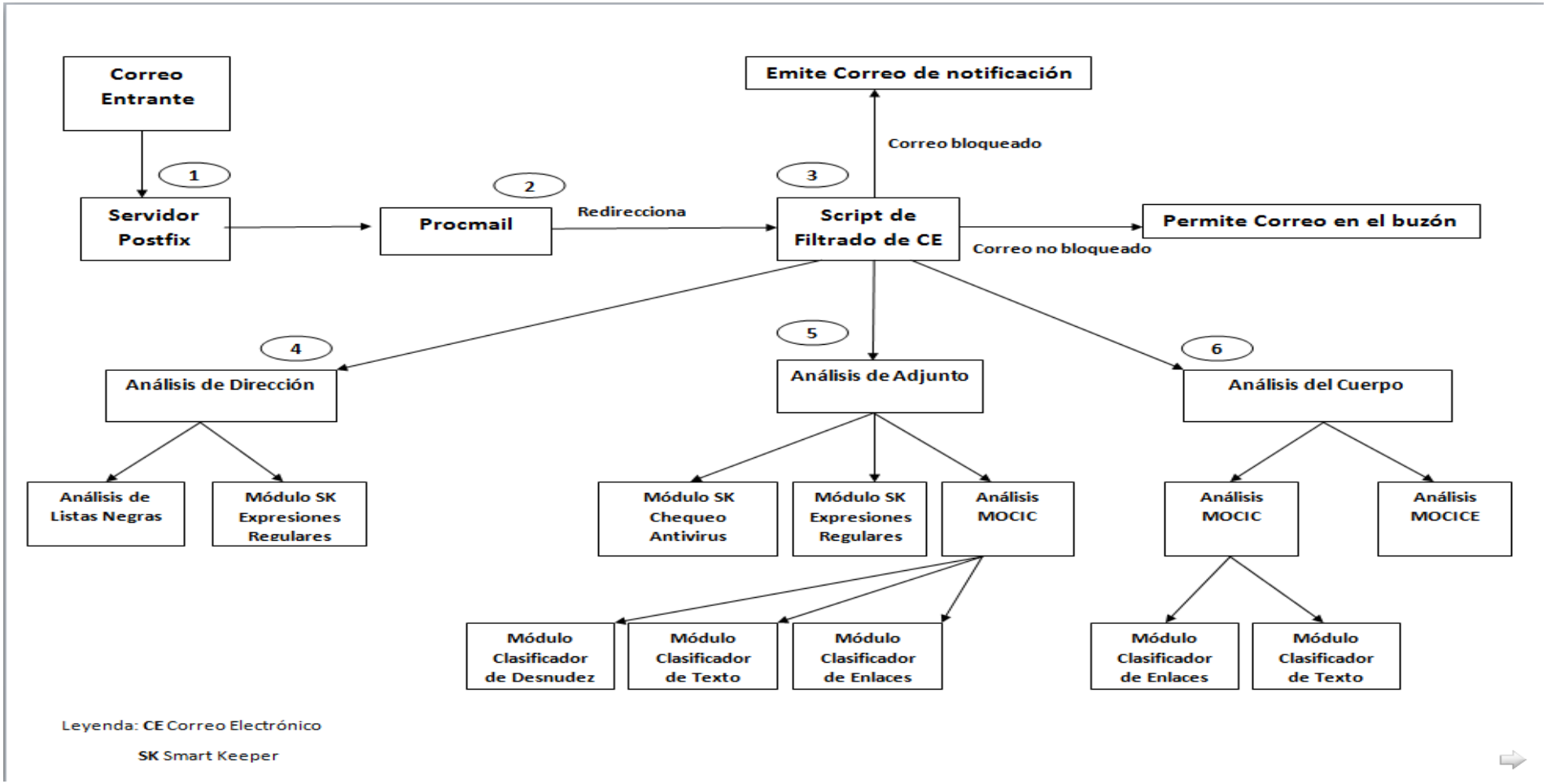
El propósito de esta encuesta es determinar los conocimientos que usted posee acerca de ciertas materias. Marque con una X de acuerdo a su criterio. **Nota:** C1 es mayor que C5.

No.	Criterio	Autoevaluación del Experto				
		C1	C2	C3	C4	C5
1	Evalúe sus conocimientos acerca de administración de servidores de correo electrónico.					
2	Evalúe sus conocimientos acerca de los sistemas de filtrado de contenido web.					
3	Evalúe sus conocimientos acerca de los sistemas de filtrado de correo electrónico.					

3. Gráfica general de validación de expertos



A continuación se muestra la Figura 2:


Correo Entrante
1
Servidor Postfix
2
Procmail
Redirecciona
3
Script de Filtrado de CE
Emite Correo de notificación
Correo bloqueado
Permite Correo en el buzón
Correo no bloqueado
4
Análisis de Dirección
Análisis de Listas Negras
Módulo SK Expresiones Regulares
5
Análisis de Adjunto
Módulo SK Chequeo Antivirus
Módulo SK Expresiones Regulares
Análisis MOCIC
Módulo Clasificador de Desnudez
Módulo Clasificador de Texto
Módulo Clasificador de Enlaces
6
Análisis del Cuerpo
Análisis MOCIC
Análisis MOCICE
Módulo Clasificador de Enlaces
Módulo Clasificador de Texto
Leyenda: CE Correo Electrónico
SK Smart Keeper