(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 732 824**

(51) Int. CI.:

H04L 12/58 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **05.09.2012 PCT/RO2012/000023**

(87) Fecha y número de publicación internacional: **01.08.2013 WO13112062**

(96) Fecha de presentación y número de la solicitud europea: **05.09.2012 E 12832751 (7)**

(97) Fecha y número de publicación de la concesión europea: **03.04.2019 EP 2807802**

(54) Título: **Sistemas y procedimientos para la detección de spam utilizando histogramas de caracteres**

(30) Prioridad:

25.01.2012 US 201213358358

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
26.11.2019

(73) Titular/es:

BITDEFENDER IPR MANAGEMENT LTD. (100.0%)
Kreontos 12
1076 Nicosia , CY

(72) Inventor/es:

DICHIU, DANIEL y
LUPESCU Z. LUCIAN

(74) Agente/Representante:

ELZABURU, S.L.P

DESCRIPCIÓN

Sistemas y procedimientos para la detección de spam utilizando histogramas de caracteres

Antecedentes

La invención se refiere a procedimientos y sistemas para clasificar comunicaciones electrónicas, y en particular a sistemas y procedimientos para filtrar comunicaciones electrónicas comerciales no solicitadas (spam).

Las comunicaciones electrónicas comerciales no solicitadas, también conocidas como spam, forman una parte importante de todo el tráfico de comunicaciones en todo el mundo, que afecta tanto a los servicios de mensajería por computadora como por teléfono. El spam puede adoptar muchas formas, desde comunicaciones por correo electrónico no solicitadas, hasta mensajes de spam que se hacen pasar por comentarios de los usuarios en diversos sitios de Internet, tales como blogs y sitios de redes sociales. El spam consume recursos de hardware valiosos, afecta la productividad y muchos usuarios de servicios de comunicación y/o Internet lo consideran molesto e intrusivo.

En el caso del correo electrónico no deseado, el software que se ejecuta en el sistema informático de un usuario o proveedor de servicios de correo electrónico se puede usar para clasificar los mensajes de correo electrónico como spam o no spam, e incluso para discriminar entre varios tipos de mensajes de spam (por ejemplo, ofertas de productos, contenido para adultos, estafas por correo electrónico). Los mensajes de spam pueden dirigirse a carpetas especiales o eliminarse.

De manera similar, el software que se ejecuta en los sistemas informáticos de un proveedor de contenido puede usarse para interceptar mensajes fraudulentos publicados en un sitio web y evitar que se muestren los mensajes respectivos, o para mostrar una advertencia a los usuarios del sitio web de que los mensajes respectivos pueden ser spam.

Se han propuesto varios procedimientos para identificar mensajes de spam, incluida la coincidencia de la dirección de origen del mensaje con las listas de direcciones de infracción o de confianza conocidas (técnicas denominadas listas negras y blancas, respectivamente), búsqueda de ciertas palabras o patrones de palabras (por ejemplo, refinanciamiento, Viagra®, stock), y análisis de encabezados de mensajes. Los procedimientos de extracción/coincidencia de características a menudo se usan junto con procedimientos de clasificación de datos automatizados (por ejemplo, filtrado bayesiano, redes neuronales).

El documento US 8001 195 B1 revela un procedimiento para la identificación de spam utilizando un algoritmo basado en histogramas y vectores léxicos (algoritmo de una pasada).

El spam a menudo llega en una sucesión rápida de grupos de mensajes similares, también conocidos como olas de spam. La forma y el contenido del spam pueden cambiar sustancialmente de una ola de spam a otra, por lo que la detección exitosa puede beneficiarse de los procedimientos y sistemas capaces de reconocer y reaccionar rápidamente a las nuevas olas de spam.

Compendio

De acuerdo con un aspecto, un procedimiento en un servidor de contenido que comprende las características de la reivindicación 13.

De acuerdo con otro aspecto, un sistema informático que comprende las características de la reivindicación 1.

De acuerdo con otro aspecto, un procedimiento en un servidor anti-spam que comprende las características de la reivindicación 14.

Breve descripción de los dibujos

Los aspectos y ventajas anteriores de la presente invención se entenderán mejor después de leer la siguiente descripción detallada y en referencia a los dibujos donde:

La Figura **1** muestra un sistema anti-spam ejemplar de acuerdo con algunas realizaciones de la presente invención.

La Figura **2** muestra una configuración de hardware ejemplar de un sistema informático servidor de acuerdo con algunas realizaciones de la presente invención.

La Figura **3-A** ilustra una transacción de detección de spam ejemplar entre una computadora cliente y un servidor anti-spam, de acuerdo con algunas realizaciones de la presente invención.

La Figura **3-B** ilustra una transacción de detección de spam ejemplar entre un servidor de contenido y un servidor anti-spam, de acuerdo con algunas realizaciones de la presente invención.

La Figura **4** muestra un indicador de objetivo ejemplar de una comunicación de destino, el indicador que comprende una cadena de destino y datos adicionales de identificación de spam, de acuerdo con algunas realizaciones de la presente invención.

La Figura **5** muestra un diagrama de un conjunto ejemplar de aplicaciones que se ejecutan en un servidor anti-spam de acuerdo con algunas realizaciones de la presente invención.

La Figura **6** muestra un histograma de caracteres ejemplar asociado a una cadena de destino y computado para una pluralidad de clases de caracteres, de acuerdo con algunas realizaciones de la presente invención.

La Figura **7** ilustra una aplicación de detector de spam ejemplar que opera en el servidor anti-spam de la Figura **1,** de acuerdo con algunas realizaciones de la presente invención.

La Figura **8** ilustra una pluralidad de grupos, cada grupo que comprende una colección de elementos similares, representados en un hiperespacio de características de acuerdo con algunas realizaciones de la presente invención.

La Figura **9** muestra una secuencia ejemplar de pasos realizados por el detector de spam de la Figura **7** de acuerdo con algunas realizaciones de la presente invención.

La Figura **10-A** muestra un resultado de un experimento de computadora, que comprende determinar el tiempo de computación para generar histogramas de caracteres para una colección de cadenas de prueba, como una función del número de caracteres distintos de las cadenas de prueba.

La Figura **10-B** muestra un resultado de un experimento de computadora, que comprende determinar el tiempo de computación para generar histogramas de caracteres para una colección de cadenas de prueba, como una función de la longitud de la cadena de las cadenas de prueba. La Figura **10-C** muestra un resultado de un experimento de computadora, que comprende determinar el tiempo de computación para calcular un conjunto de distancias entre cadenas para una colección de cadenas de prueba, como una función del número de caracteres distintos de las cadenas de prueba.

La Figura **10-D** muestra un resultado de un experimento de computadora, que comprende determinar el tiempo de computación para calcular un conjunto de distancias entre cadenas para una colección de cadenas de prueba, en función de la longitud de la cadena de las cadenas de prueba.

La Figura **11** muestra una gráfica de longitud de cadena frente a un indicador de marca de tiempo, para una colección de comentarios reales de blog, que incluyen tanto spam como no spam.

La Figura **12** muestra una gráfica de una serie de caracteres distintos frente a un indicador de marca de tiempo, para una colección de comentarios de blog reales, que incluyen tanto spam como no spam.

Descripción detallada de realizaciones preferidas

En la siguiente descripción, se entiende que todas las conexiones citadas entre estructuras pueden ser conexiones operativas directas o conexiones operativas indirectas a través de estructuras intermedias. Un conjunto de elementos incluye uno o más elementos. Se entiende que cualquier recitación de un elemento se refiere a al menos un elemento. Una pluralidad de elementos incluye al menos dos elementos. A menos que se requiera lo contrario, no es necesario que los pasos de los procedimientos descritos se realicen en un orden ilustrado particular. Un primer elemento (por ejemplo, datos) derivado de un segundo elemento abarca un primer elemento igual al segundo elemento, así como un primer elemento generado al procesar el segundo elemento y opcionalmente otros datos. La toma de una determinación o decisión según un parámetro abarca la determinación o decisión según el parámetro y, opcionalmente, según otros datos. A menos que se especifique lo contrario, un indicador de cierta cantidad/datos puede ser la cantidad/datos mismo, o un indicador diferente de la cantidad/datos mismos. Los programas de computadora descritos en algunas realizaciones de la presente invención pueden ser entidades de software independientes o subentidades (por ejemplo, subrutinas, objetos de código) de otros programas de computadora. A menos que se especifique lo contrario, el término spam no se limita a correo no deseado, sino que también abarca comunicaciones electrónicas, tales como contenido generado por usuario comercial no legítimo o no solicitado, en forma de comentarios de blog, discusiones de foros, entradas de wiki, comentarios de clientes, publicaciones en sitios de redes sociales, mensajes instantáneos, así como mensajes de texto telefónico y multimedia, entre otros. Los medios legibles por computadora abarcan medios de almacenamiento no transitorios, tales como medios magnéticos, ópticos y semiconductores (por ejemplo, discos duros, discos ópticos, memoria flash, DRAM), así como enlaces de comunicaciones como cables conductores y enlaces de fibra óptica. De acuerdo con algunas realizaciones, la presente invención proporciona, *Entre otros,* sistemas informáticos que comprenden hardware programado para realizar los procedimientos descritos en este documento, así como medios legibles por computadora que codifican instrucciones para realizar los procedimientos descritos en este documento.

La siguiente descripción ilustra realizaciones de la invención a modo de ejemplo y no necesariamente a modo de limitación.

La Figura **1** muestra un sistema anti-spam ejemplar de acuerdo con algunas realizaciones de la presente invención. El sistema **10** Incluye una pluralidad de servidores de contenido. **12a-b,** un servidor anti-spam **16,** y una pluralidad de sistemas cliente **14a-b.** Los servidores de contenido **12a-b** pueden representar a los servidores web que alojan y/o entregan contenido en línea, tales como sitios web personales y corporativos, blogs, sitios de redes sociales y sitios

de entretenimiento en línea, entre otros. Otros servidores de contenido **12a-b** pueden representar a los servidores de correo electrónico que proporcionan la entrega de mensajes electrónicos a los sistemas cliente. **14a-b. Los s**istemas cliente **14a-b** pueden representar computadoras de usuario final, cada una con procesador, memoria y almacenamiento, y ejecutar un sistema operativo tal como Windows®, MacOS® o Linux. Algunos sistemas informáticos cliente **14a-b** pueden representar dispositivos móviles de computación y/o telecomunicaciones, tales como PC tabletas, teléfonos móviles y asistentes digitales personales (PDA). En algunas realizaciones, los sistemas cliente **14a-b** pueden representar a clientes individuales, o varios sistemas cliente pueden pertenecer al mismo cliente. El servidor anti-spam **16** puede incluir uno o más sistemas informáticos. Una red **18** conecta servidores de contenido **12a-b,** sistemas cliente **14a-b,** y servidor anti-spam **16.** La red **18** puede ser una red de área amplia tal como Internet, mientras que partes de la red **18** también pueden incluir una red de área local (LAN).

La Figura **2** muestra una configuración de hardware ejemplar de un sistema informático servidor, tal como un servidor anti-spam **16.** En algunas realizaciones, el servidor **16** comprende un procesador **20,** una unidad de memoria **22,** un conjunto de dispositivos de almacenamiento **24,** y un controlador de interfaz de comunicación **26,** todos conectados por un conjunto de buses **28.**

En algunas realizaciones, el procesador **20** comprende un dispositivo físico (por ejemplo, un circuito integrado de múltiples núcleo) configurado para ejecutar operaciones computacionales y/o lógicas con un conjunto de señales y/o datos. En algunas realizaciones, tales operaciones lógicas se entregan al procesador **20** en forma de una secuencia de instrucciones del procesador (por ejemplo, código de máquina u otro tipo de software). La unidad de memoria **22** puede comprender medios legibles por computadora (por ejemplo, RAM) que almacenan datos/señales accedidas o generadas por el procesador **20** en el curso de llevar a cabo las instrucciones. Los dispositivos de almacenamiento **24** incluyen medios legibles por computadora que permiten el almacenamiento no volátil, la lectura y la escritura de instrucciones y/o datos de software. Los dispositivos de almacenamiento ejemplares **24** incluyen discos magnéticos y ópticos y dispositivos de memoria de semiconductores (por ejemplo, flash), así como medios extraíbles tales como discos y unidades de CD y/o DVD. El controlador de interfaz de comunicación **26** habilita al sistema del servidor **16** para conectarse a la red **18** y/o a otras máquinas/sistemas informáticos. Los controladores de interfaz de comunicación típicos **26** incluyen adaptadores de red. Los buses **28** representan colectivamente la pluralidad de buses de sistema, periféricos y chipset, y/o todos los otros circuitos que permiten la intercomunicación de los dispositivos **20-26** del sistema servidor **16.** Por ejemplo, los buses **28** pueden comprender el procesador de conexión de bus northbridge **20** a la memoria **22,** y/o el procesador de conexión de bus southbridge **20** a dispositivos **24-26,** entre otros.

En algunas realizaciones, cada sistema cliente **14a-b** comprende una aplicación de lector de documentos (por ejemplo, navegador web, lector de correo electrónico, reproductor de medios), que puede ser un programa de computadora utilizado para acceder de forma remota a los datos almacenados en servidores de contenido **12a-b.** Cuando un usuario accede a un documento en línea, tal como una página web, o recibe una comunicación electrónica tal como un correo electrónico, los datos asociados al documento/comunicación circulan en partes de la red **18** entre el respectivo servidor de contenido y el sistema cliente **14.** En algunas realizaciones, la aplicación del lector recibe los datos del documento, los traduce a forma visual y los muestra al usuario. Algunas realizaciones de la aplicación del lector también pueden permitir que el usuario interactúe con el contenido mostrado. En el caso de correo electrónico, el sistema cliente. **14a-b** puede incluir software especializado configurado para clasificar el correo electrónico entrante en una de una pluralidad de categorías (por ejemplo, spam, legítimo, varias otras clases y subclases).

En algunas realizaciones, los servidores de contenido **12a-b** están configurados para recibir contenido generado por el usuario (por ejemplo, artículos, entradas de blog, cargas de medios, comentarios, etc.) de una pluralidad de usuarios, y para organizar, formatear y entregar dicho contenido a terceros a través de la red **18.** Una parte de los datos generados por el usuario recibidos en los servidores **12a-b** puede comprender comunicaciones electrónicas que pueden incluir mensajes no solicitados (spam). Ejemplos de dichas comunicaciones electrónicas, denominadas comunicaciones de destino o mensajes de destino en la siguiente descripción, son mensajes de correo electrónico, comentarios de blogs, publicaciones en redes sociales y comentarios enviados a sitios web de entretenimiento y/o noticias, entre otros.

En algunas realizaciones, los servidores de contenido **12a-b** pueden comprender componentes de software configurados para procesar las comunicaciones de destino recibidas de los usuarios para detectar el spam. Cuando se detectan, los mensajes de spam (por ejemplo, comentarios de blog fraudulentos, etc.) pueden bloquearse y/o evitarse que se muestren dentro del sitio web respectivo. El software de procesamiento de spam puede implementarse en servidores de contenido **12a-b** en la forma de scripts del lado del servidor. Dichos scripts pueden incorporarse como complementos en paquetes de scripts más grandes, por ejemplo. como complemento anti-spam para las plataformas de publicación en línea de Wordpress® o Drupal®. En algunas realizaciones, para detectar spam, servidores. **12a-b** puede configurarse para participar en una transacción de detección de spam colaborativa con un servidor anti-spam **16,** como se describe en detalle a continuación.

En algunas realizaciones, el servidor anti-spam **16** está configurado para realizar una transacción colaborativa de detección de spam con servidores de contenido **12a-b** y/o sistemas cliente **14a-b.** La Figura **3-A** ilustra un intercambio de datos ejemplar entre el servidor **16** y un sistema cliente **14,** en una realización configurada para detectar spam de correo electrónico. El sistema cliente **14** puede recibir mensajes de correo electrónico de un servidor de correo a través de la red. **18.** Después de recibir un mensaje de correo electrónico, el sistema cliente **14** puede enviar un indicador de

destino **40** que comprende datos de identificación de spam asociados al mensaje de correo electrónico al servidor anti-spam **16,** y puede recibir en respuesta una etiqueta de destino **50** indicativa de si el mensaje de correo electrónico respectivo es spam. El sistema cliente **14** posteriormente, puede colocar el mensaje en una categoría de mensaje apropiada (por ejemplo, spam o correo electrónico legítimo).

La Figura **3-B** ilustra un intercambio de datos ejemplar entre el servidor anti-spam **16** y un servidor de contenido **12,** en una realización configurada para detectar el spam generado por el usuario recibido en el servidor **12,** tal como el spam de blog o spam publicado en las redes sociales, entre otros. Después de recibir una comunicación de destino, por ejemplo, como resultado de un usuario que publica un comentario en un sitio web alojado en el servidor **12,** el servidor **12** puede enviar indicador de destino **40** que comprende datos de identificación de spam extraídos de la comunicación de destino al servidor anti-spam **16,** y en respuesta recibir la etiqueta de destino **50** que indica si la comunicación de destino es spam.

La Figura **4** muestra un indicador de objetivo ejemplar de una comunicación de destino, de acuerdo con algunas realizaciones de la presente invención. El indicador de destino **40** comprende un ID de objeto **41** (por ejemplo, etiqueta, hash) que identifica de manera única la comunicación de destino y una cadena de destino **42** incluida una parte de texto de la comunicación de destino, tal como una parte de texto de un comentario de blog. En algunas realizaciones, la cadena de destino **42** comprende sustancialmente todo el texto de la comunicación de destino. Cuando la comunicación de destino incluye varias partes de texto, la cadena de destino **42** puede comprender una concatenación de partes de texto; alternativamente, cada parte de texto puede recibir una cadena de destino distinta **42.** En algunas realizaciones, la cadena de destino **42** comprende una sección de la parte de texto de la comunicación de destino, la sección que tiene una longitud de cadena preestablecida (por ejemplo, 128 caracteres consecutivos).

Algunas realizaciones del indicador de destino **40** pueden incluir otros datos de identificación de spam correspondientes a la comunicación de destino junto a la cadena de destino **42.** En el ejemplo de la Figura **4,** el indicador de destino **40** incluye un indicador de nombre de usuario **44** indicativo de un nombre (por ejemplo, nombre personal, seudónimo, nombre de pantalla, nombre de inicio de sesión, avatar, identificador, etc.) proporcionado por el remitente de la comunicación de destino, un indicador de dirección **46** indicativo de una ubicación de origen (por ejemplo, una dirección IP de origen) de la comunicación de destino, y una marca de tiempo **48** es indicativo de un punto en tiempo real (por ejemplo, fecha y hora) en que se envió la comunicación de destino.

En algunas realizaciones, la etiqueta de destino **50** puede comprender un ID de objeto tal como ID **41,** y un indicador del estado de spam de la comunicación de destino. La etiqueta de destino **50** especifica efectivamente si la comunicación de destino es spam, según la evaluación realizada por el servidor anti-spam **16,** como se describe en detalle a continuación.

La Figura **5** muestra un diagrama de un conjunto ejemplar de aplicaciones que se ejecutan en un servidor anti-spam **16** de acuerdo con algunas realizaciones de la presente invención. Las aplicaciones incluyen un detector de spam **32** y un administrador de comunicación **34** conectado o detector de spam **32.** Las aplicaciones **32** y **34** pueden ser programas de computadora independientes o pueden formar parte de conjuntos de software más grandes que brindan, por ejemplo, servicios de seguridad informática tal como la detección de malware. En algunas realizaciones, el servidor anti-spam **16** también alberga una base de datos anti-spam **30.** Alternativamente, la base de datos anti-spam puede residir en un sistema informático distinto del servidor **16,** pero conectado al servidor **16** a través de la red **18,** o en medios legibles por computadora conectados al servidor **16.**

La base de datos anti-spam **30** comprende un repositorio de conocimientos relacionados con el spam en línea. En algunas realizaciones, la base de datos **30** comprende un corpus de histogramas de caracteres, cada histograma calculado para una comunicación electrónica como se describe más adelante. La base de datos **30** puede incluir datos tanto para comunicaciones electrónicas spam como no spam, tales como comentarios de blogs, comentarios publicados en sitios de redes sociales, etc. Además de los datos de histogramas, cada registro almacenado en la base de datos **30** puede incluir información adicional, tal como cadena **42** de la comunicación de destino respectiva, un indicador de tiempo tal como marca de tiempo **48,** e indicadores de longitud de cadena, número de caracteres distintos y puntuación de cadena de la cadena **42;** los usos ejemplares de tales cantidades se describen a continuación. En algunas realizaciones, en relación con cada histograma de caracteres, la base de datos **30** puede almacenar un indicador de asignación de grupo indicativo de un grupo de mensajes a los que se asigna actualmente la cadena correspondiente. La base de datos **30** también puede almacenar una estructura de datos que comprende una pluralidad de identificadores, tal como el ID de objeto **41,** cada identificador de objeto asociado de manera única a una comunicación electrónica, y un mapeo que asocia cada histograma de caracteres con la comunicación de destino para la que fue calculado, lo que permite el detector de spam **32** recuperar selectivamente histogramas de la base de datos **30,** como se muestra abajo.

La Figura **6** muestra un histograma de caracteres ejemplar calculado para la cadena de destino **42** de acuerdo con algunas realizaciones de la presente invención. El histograma de destino **60** comprende un conjunto de números, cada número indica un recuento de ocurrencias de cada carácter distinto dentro de la cadena de destino **42.** Por ejemplo, un histograma de destino de "Mississippi" incluye 1 para "M", 4 para "i", 4 para "s" y 2 para "p". En algunas realizaciones, los caracteres se agrupan en varias clases de caracteres distintos **62,** como "minúsculas", "mayúsculas",

"dígitos" y "caracteres especiales", entre otros. El número y la composición de varias clases de caracteres pueden variar entre las realizaciones.

En algunas realizaciones, el administrador de comunicación. **34** está configurado para gestionar la comunicación con los sistemas cliente. **14a-b** y/o servidores de contenido **12a-b.** Por ejemplo, el administrador **34** puede establecer conexiones a través de la red **18,** enviar y recibir datos tales como indicadores de destino y etiquetas de destino a/desde los sistemas cliente **14a-b,** y servidores de contenido **12a-b.**

La Figura **7** muestra un diagrama de un detector de spam ejemplar que opera en un servidor anti-spam **16** de acuerdo con algunas realizaciones de la presente invención. El detector de spam **32** comprende un administrador de histograma **36** y un motor de agrupamiento **38** conectado al administrador de histogramas **36.** El detector de spam **32** puede recibir el indicador de destino **40** de una comunicación de destino del administrador de comunicación **34** (ver figs. **5** y **3-A-B**) y un histograma de referencia **64** de la base de datos anti-spam **30,** y a su vez da salida a la etiqueta de destino **50** al administrador de comunicación **34,** para ser reenviada al sistema cliente **14** o servidor de contenido **12** que inició la respectiva transacción de detección de spam.

En algunas realizaciones, el administrador de histogramas. **36** está configurado para recibir el indicador de destino **40** del administrador de comunicación **34,** para calcular el histograma de destino **60** a partir de datos del indicador de destino **40,** para realizar un proceso de filtrado previo para determinar un conjunto de histogramas de referencia elegibles **64,** para recuperar selectivamente los histogramas de referencia **64** de la base de datos anti-spam **30,** y para remitir histogramas. **60** y **64** al motor de agrupamiento **38** para la comparación de histogramas y asignación de grupos. El funcionamiento del administrador de histogramas. **36** será discutido en detalle en relación con la Figura **9.**

En algunas realizaciones, el análisis del mensaje de destino se realiza en un hiperespacio de la característica de mensaje, al analizar las distancias entre un vector de característica correspondiente a una comunicación de destino y un conjunto de vectores representativos, cada uno de los cuales define una colección distinta (grupo) de mensajes. La Figura **8** muestra tres grupos de mensajes ejemplares **70a-c** formados por vectores de características **74a-c,** respectivamente, en un espacio de características 2-D simple que tiene dos ejes, d1 y d2. En algunas realizaciones de la presente invención, los ejes d1 y d2 corresponden a características distintas del histograma de caracteres. Por ejemplo, el eje d1 puede corresponder al carácter "M" y el eje d2 al carácter "s". Luego, la cadena "Mississippi" puede representarse por el vector de característica (1,4), considerando que "M" aparece una vez, mientras que "s" aparece cuatro veces dentro de la cadena correspondiente. Dos cadenas que tienen histogramas de caracteres similares se encuentran cerca una de la otra en este hiperespacio de características ejemplar: en el ejemplo anterior, "Mississippi" y "Misión: imposible" tienen vectores de características idénticos en este espacio. En algunas realizaciones, cada grupo de mensajes **70** consiste en mensajes que ocupan una región sustancialmente pequeña de hiperespacio de características, lo que significa que todos los miembros de un grupo tienen vectores de características similares, es decir, histogramas de caracteres similares.

En algunas realizaciones, el motor de agrupamiento **38** está configurado para mantener una colección de grupos de grupo mensajes **70,** lo que representa un corpus de comunicaciones recibidas en el servidor anti-spam. **16** durante un historial de acumulación de datos, agrupados en grupos según la similitud. Algunos grupos **70** pueden representar olas de spam individuales, cada una de las cuales incluye copias o variantes del mismo mensaje de spam enviado a una multitud de clientes y/o publicado en una multitud de sitios web. Idealmente, un grupo de mensajes consiste en cadenas de destino idénticas o casi idénticas. El motor de agrupamiento **38** además está configurado para recibir el histograma de destino **60** y asignar la comunicación de destino representada por histograma. **60** a un grupo de mensajes que es más similar según la similitud del histograma. Para realizar la asignación del grupo, el motor de agrupamiento **38** recibe un conjunto de histogramas de referencia **64** del administrador de histogramas **36,** cada histograma **64** representativo de un grupo, y compara histogramas **60** y **64** para determinar qué grupo de mensajes coincide mejor con el histograma de destino **60.** Más detalles del funcionamiento del motor de agrupamiento. **38** se dan a continuación, en relación a la Figura **9.**

La Figura **9** muestra una secuencia ejemplar de pasos realizados por el detector de spam **32** (Figura **7**) dentro de una transacción de detección de spam, de acuerdo con algunas realizaciones de la presente invención. En un paso **102,** el detector de spam **32** recibe el indicador de destino **40** del sistema cliente **14** o servidor de contenido **12,** a través del administrador de comunicación **34.** A continuación, en un paso. **104,** el detector de spam extrae la cadena de destino **42** según indicador de destino **40,** y calcula el histograma de caracteres **60** de la cadena **42.** El paso **104** también puede comprender la computación de varios parámetros de la cadena de destino **42,** tal como la longitud de la cadena y/o el número de caracteres distintos, que se utilizan para filtrar previamente la colección de grupos de mensajes en un paso **106.** En algunas realizaciones, en el paso **106,** el detector de spam **32** realiza una operación de prefiltración para seleccionar, de acuerdo con un conjunto de condiciones de prefiltración, un subconjunto de agrupaciones de mensajes candidatos de la colección completa mantenida por el motor de agrupamiento **38.** Al seleccionar solo un (pequeño) subconjunto de grupos con los cuales realizar comparaciones de histogramas, el detector de spam **32** puede reducir efectivamente los costos computacionales.

En algunas realizaciones, el detector de spam. **32** puede seleccionar el subconjunto de grupos candidatos según la longitud de la cadena. La longitud de la cadena de la cadena de destino **42** se compara con la longitud de cadena de un representante de cada grupo, o con una longitud de cadena promedio de los miembros del grupo respectivo. Un

grupo puede seleccionarse como candidato para la comparación de histogramas cuando su longitud de cadena típica está dentro de un umbral predeterminado de la longitud de cadena de la cadena de destino **42.**

Un criterio alternativo de prefiltraciones el número (recuento) de caracteres distintos. Por ejemplo: la cadena "Mississippi" tiene 4 caracteres distintos: M, i, s y p. Para cada grupo, el número de caracteres distintos de la cadena de destino **42** se compara con el número de caracteres distintos de un miembro representativo del grupo respectivo, o con un número promedio de caracteres distintos de los miembros de un grupo; los grupos que tienen números similares de caracteres distintos como cadena de destino **42** se seleccionan como candidatos para la comparación del histograma.

En algunas realizaciones, la prefiltración puede proceder de acuerdo con una puntuación de cadena calculada de la siguiente manera:

$$S = \sum_i p_i w_i , \qquad [1]$$

donde *i* indexa los caracteres de la cadena, p_i denota un indicador de posición de carácter *i* dentro del conjunto de todos los caracteres (por ejemplo, un código ASCII del carácter respectivo), y w_i denota un peso específico del carácter del carácter respectivo. En algunas realizaciones, los caracteres se dividen en varias clases, tales como las clases **62** Ilustradas en la Figura 6: minúsculas, mayúsculas, dígitos y caracteres especiales, entre otros. Las ponderaciones *w(i)* pueden ser idénticas dentro de la misma clase de caracteres, pero pueden diferir de una clase a otra. Por ejemplo, la ponderación asociada a un carácter especial puede ser mayor que la ponderación de una letra minúscula. Para cada grupo, la puntuación de la cadena [1] de la cadena de destino **42** se compara con la puntuación de cadena de un miembro representativo del grupo respectivo o con una puntuación de cadena promedio del grupo; los grupos que tienen puntuaciones de cadena similares a la cadena de destino **42** sSe seleccionan como candidatos para la comparación del histograma.

En algunas realizaciones, los criterios de prefiltración pueden combinarse. Por ejemplo, un primer subconjunto de grupos de mensajes puede seleccionarse según la similitud de la puntuación de la cadena; luego, fuera del primer subconjunto de agrupaciones, se selecciona un segundo subconjunto, en el que cada grupo tiene una longitud de cadena similar y un número similar de caracteres distintos a la cadena de destino **42.**

Habiendo seleccionado un conjunto de candidatos para la comparación de histogramas, para cada grupo seleccionado, el detector de spam **32** puede instruir al administrador de histogramas **36** para recuperar selectivamente de la base de datos anti-spam **30** un histograma de referencia **64** que corresponde a un mensaje representativo del grupo respectivo. Entonces, una secuencia de pasos de bucle **108-116** se ejecuta para cada grupo seleccionado. En un paso **108,** el detector de spam puede verificar si todos los grupos seleccionados fueron evaluados por comparación de histograma. Si es así, el detector de spam **32** procede a un paso **118** descrito abajo. Si es no, en un paso **110,** se evalúa el siguiente grupo. En un paso **112,** el histograma de destino **60** se compara con el histograma de referencia **64.**

En algunas realizaciones, el paso **112** comprende el cálculo de un conjunto de distancias entre cadenas indicativas del grado de similitud entre los histogramas **60** y **64.** En algunas realizaciones, una distancia entre cadenas entre dos cadenas s_1 y s_2 puede formularse como:

$$D_1(s_1, s_2) = \sum_{i \in s_1 \cap s_2} w_i \left| N_1^i - N_2^i \right| , \qquad [2]$$

donde *i* indexa el subconjunto de caracteres comunes a la cadena s_1 y s_2, w_i es la ponderación del carácter *i*, N_1^i denota el recuento de ocurrencias del carácter *i* dentro de la cadena s_1 y en donde N_2^i denota el recuento de ocurrencias de carácter *i* dentro de la cadena s_2. Una distancia entre cadenas alternativa está dada por:

$$D_2(s_1, s_2) = \sum_{i \in s_1 - s_2} w_i \cdot c , \qquad [3]$$

donde *i* indexa el subconjunto de caracteres presentes solo en s_1, pero no en s_2, w_i es la ponderación del carácter *i* y *c* es una constante predeterminada, independiente del carácter. Como se discutió anteriormente, las ponderaciones w_i pueden ser específicas de un carácter o específicas de una clase (por ejemplo, caracteres especiales versus. letras minúsculas). Una razón para usar ponderaciones específicas de los caracteres es que algunos caracteres se usan con más frecuencia que otros para la ofuscación de texto, un procedimiento empleado frecuentemente por los spammers y que consiste en reemplazar ciertos caracteres en un texto con otros caracteres (por ejemplo, "vi4gra"), para evitar la detección del spam.. Al asignar ponderaciones relativamente pequeñas a los caracteres que se utilizan en la ofuscación, las versiones ofuscadas de una cadena de destino pueden parecer muy similares entre sí según la distancia entre cadenas y, por lo tanto, todas pueden identificarse correctamente como spam. El valor de *c* puede ser utilizado como un parámetro de ajuste: si *c* es demasiado pequeño, dos cadenas bastante diferentes pueden

considerarse erróneamente similares; si *c* es demasiado grande, pequeñas diferencias entre las cadenas pueden ser excesivamente amplificadas.

Algunas realizaciones pueden calcular una distancia entre cadenas combinada:

$$D_3(s_1, s_2) = D_1(s_1, s_2) + D_2(s_1, s_2). \quad [4]$$

Además, debido a que D_2 no es conmutativa, una distancia entre cadenas alternativa es:

$$D_4(s_1, s_2) = D_2(s_1, s_2) + D_2(s_2, s_1). \quad [5]$$

En algunas realizaciones, el paso **112** (Figura **9**) comprende la computación $D_1(T, R)$ y/o $D_2(T, R)$, donde *T* denota cadena de destino **42** y *R* denota la cadena de referencia asociada al histograma de referencia **64.** Alternativamente, el detector de spam puede calcular $D_3(T, R)$, $D_3(R, T)$, y/o $D_4(T, R)$. A continuación, un paso. **114** determina si se encuentra una coincidencia entre el histograma de destino **60** e histograma de referencia **64.**

En algunas realizaciones, una coincidencia de histograma requiere que una distancia entre cadenas sea menor que un umbral predeterminado. Por ejemplo, una coincidencia de histograma puede requerir que *o bien* $D_1<t_1$ *o* $D_1<t_2$, o eso *ambos* D_1 y D_2 sean menores que sus respectivos umbrales. Alternativamente, una coincidencia de histograma requiere que $D_3 = D_1 + D_2 < t_3$. En otra realización más, tanto $D_3(T, R)$ *como* $D_3(R, T)$ deben ser menores que un umbral, o $D_4 <t_4$ para una coincidencia exitosa. Los valores umbral t_i pueden ser independientes de las cadenas que se comparan, o pueden variar según la longitud de la cadena y/o el número de caracteres distintos de la cadena de destino **42.** En algunas realizaciones, los valores de umbral más altos se utilizan para cadenas comparativamente más largas, o cadenas con un número comparativamente mayor de caracteres distintos.

Cuando una coincidencia entre histogramas **60** y **64** se encuentra, el detector de spam **32** procede a un paso **116,** en el que el grupo de mensajes respectivo está marcado como elegible para recibir la cadena de destino **42.** Si los histogramas no coinciden, el detector de spam **32** vuelve al paso **108.** Cuando todos los grupos de mensajes seleccionados en el paso **106** han sido evaluados por comparación de histograma, un paso **118** determina si algún grupo es elegible para recibir la cadena de destino **42.** Si es así, en un paso **122** el detector de spam **32** puede calcular, para cada grupo elegible, un indicador de similitud de cadena a grupo indicativo de qué tan similar es la cadena de destino **42** es a todos los miembros del grupo respectivo. Un indicador de similitud de cadena a grupo ejemplar de cada grupo elegible comprende la fracción de miembros del grupo que tienen histogramas de referencia que coinciden con el histograma de destino **60.**

A continuación, un paso. **124** lleva a cabo la asignación real de la cadena de destino **42** al grupo de mensajes a la que es más similar. En algunas realizaciones, la cadena de destino **42** se asigna al grupo con el indicador similitud de cadena a grupo mayor, determinado en el paso **122.** El motor de agrupamiento **38** puede actualizar los datos de asignación del grupo para reflejar la adición de un nuevo miembro del grupo y un registro de la cadena de destino **42** puede ser introducido en la base de datos anti-spam **30.** En algunas realizaciones, el paso **124** Incluye además la determinación de un conjunto de parámetros de identificación de spam asociados al grupo que recibe el mensaje de destino. Por ejemplo, algunas realizaciones pueden calcular un intervalo de tiempo transcurrido entre marcas de tiempo sucesivas, dentro del mismo grupo. Dichos parámetros se pueden guardar en relación con cada grupo y se pueden usar para determinar automáticamente (sin la supervisión de un operador humano) si un grupo particular incluye spam o mensajes legítimos, o si un grupo particular puede representar una ola de spam.

Cuando no se encontraron grupos elegibles para recibir el mensaje de destino (paso **118**), lo que indica que esa cadena de destino **42** es probable que sea distinta de cualquiera ya almacenada en la base de datos anti-spam, en un paso **120** el motor de agrupamiento **38** puede crear un nuevo grupo con el mensaje de destino como único miembro, y puede guardar un registro del mensaje de destino en la base de datos anti-spam **30.**

En un paso **126,** el detector de spam **32** puede determinar la etiqueta de destino **50** Identificando la comunicación de destino como spam o legítima. En algunas realizaciones, la decisión de si la comunicación de destino es spam se toma de acuerdo con la asignación del grupo de la cadena de destino **42.** Cuando la cadena **42** se asigna a un grupo que consiste principalmente en mensajes de spam, entonces el mensaje de destino también puede recibir una etiqueta de spam.

En algunas realizaciones, la etiqueta **50** se determina de acuerdo con ciertas características de identificación de spam de los miembros del grupo a los que se asignó el mensaje de destino. Una de esas características de identificación de spam es la marca de tiempo **48.** La asignación de numerosos miembros nuevos en un breve intervalo de tiempo puede ser una indicación de que el grupo respectivo consiste en una ola de mensajes de spam. En algunas realizaciones, el detector de spam. **32** puede determinar un intervalo de tiempo transcurrido entre una pluralidad de marcas de tiempo asociadas a miembros de un grupo, por ejemplo, el intervalo de tiempo más corto dentro del cual 6 miembros fueron asignados a ese grupo, y cuando el intervalo de tiempo cae por debajo de un umbral predeterminado, marca el grupo

respectivo como spam. En algunas realizaciones, el recuento de miembros del grupo se puede usar como una característica de identificación de spam: cuando un grupo adquiere un exceso de un número predeterminado de miembros, el grupo respectivo puede marcarse como spam.

Los sistemas y procedimientos ejemplares descritos anteriormente permiten que un sistema anti-spam detecte comunicaciones no solicitadas en forma de contenido generado por el usuario en Internet en forma de comentarios de blog, comentarios publicados en sitios de redes sociales, etc., y también contenido en el sitio. forma de mensajes de correo electrónico, mensajes instantáneos y mensajes de texto telefónico y multimedia.

En algunas realizaciones, un sistema informático extrae una cadena de destino de caracteres de una comunicación electrónica tal como un comentario de blog, lo transmite a un servidor de spam o no spam y recibe un indicador de si la comunicación electrónica respectiva es spam o no spam del. servidor anti-spam. Cuando la comunicación electrónica es spam, el sistema informático puede bloquear, poner en cuarentena, borrar o restringir de cualquier otra manera la visualización de la comunicación electrónica y/o puede emitir una advertencia al usuario.

El servidor anti-spam determina si la comunicación electrónica es spam o no spam según la frecuencia de aparición de ciertos caracteres dentro de la cadena de destino. Un histograma de caracteres de la cadena de destino se calcula y se compara con los histogramas calculados para un corpus de comunicaciones electrónicas, posiblemente incluyendo mensajes de spam y de no spam. Los procedimientos y sistemas descritos en la presente invención aprovechan la observación de que dos cadenas similares siempre tienen histogramas de caracteres similares. Por lo tanto, encontrar una coincidencia entre el histograma de la cadena de destino y otro histograma calculado para una cadena de referencia puede ser una indicación de que la cadena de destino es similar a la cadena de referencia. En tal caso, el servidor anti-spam puede determinar si la comunicación electrónica es spam según si la cadena de referencia es indicativa de spam, por ejemplo, si la cadena de referencia pertenece a un grupo de comunicaciones electrónicas etiquetadas como spam.

Sin embargo, hay muchas situaciones en las que dos cadenas distintas tienen histogramas muy similares. Para evitar una identificación falsa positiva, se pueden considerar otras características de la cadena de destino, tales como una marca de tiempo, al tomar una decisión si dos cadenas son similares.

La comparación de histogramas puede no ser un procedimiento confiable para identificar cadenas similares cuando la longitud de la cadena excede un cierto umbral. En el límite de cadenas muy largas, todas las cadenas tienen histogramas muy similares, que simplemente indican una frecuencia natural de ocurrencia de cada carácter en el idioma respectivo. Por lo tanto, los sistemas y procedimientos descritos aquí son particularmente adecuados para analizar cadenas cortas, que aparecen en las comunicaciones electrónicas, tales como publicaciones de blog y comentarios en sitios de redes sociales tales como Facebook® y Twitter®.

Un problema adicional para la detección de spam mediante la comparación de cadenas es la ofuscación, en la cual los spammers pueden reemplazar ciertos caracteres en un mensaje con otros caracteres (por ejemplo, Vi4gra), para evitar la detección. La ofuscación de la cadena se puede abordar empleando una ponderación específica para cada carácter y determinando una distancia entre cadenas de acuerdo con el recuento de cada carácter *y* la ponderación, como en las fórmulas [2-3]. Las ponderaciones específicas de los caracteres permiten ajustar la sensibilidad de la comparación de cadenas. En algunas realizaciones, los caracteres pueden agruparse por categorías (por ejemplo, letras, dígitos, caracteres especiales), todos los caracteres de una categoría reciben una ponderación idéntica, específica de la categoría. Dicha agrupación puede abordar la ofuscación de texto, ya que, por ejemplo, la ofuscación con letras aleatorias es más frecuente que la ofuscación de texto con otros tipos de caracteres. Cuando se usan ciertos caracteres en la ofuscación más que en otros, asignándoles una ponderación comparativamente menor reduce ventajosamente la distancia entre cadenas entre dos versiones ofuscadas de la misma cadena, haciendo que las dos cadenas parezcan más similares.

Algunas realizaciones de la presente invención organizan el corpus en una pluralidad de grupos, cada grupo de registros consiste en cadenas similares. En lugar de realizar comparaciones de histogramas en todos los registros del corpus, la agrupación permite comparar la cadena de destino con solo una cadena de destino representativa por grupo, lo que reduce significativamente los costos computacionales.

La agrupación también puede facilitar la detección automática de spam (sin supervisión). A diferencia de los procedimientos de agrupamiento convencionales, en los que la clasificación de mensajes se logra comúnmente mediante un entrenamiento supervisado de un clasificador, por ejemplo, en un cuerpo de entrenamiento previamente clasificado en una pluralidad de clases de mensajes predeterminadas, algunas realizaciones de la presente invención realizan una agrupación dinámica sin conocimiento previo del estado del spam (spam versus no spam) de grupos o mensajes. Un grupo se puede identificar automáticamente como representante de spam o cuando se acumula un cierto número de miembros en un corto intervalo de tiempo.

En un experimento de computadora, un corpus de 22,000 comentarios de blog se clasificó en grupos de acuerdo con algunas realizaciones de la presente invención. Los cálculos se realizaron en un sistema informático equipado con un procesador Pentium 4 a 3 GHz y 1,5 GB de RAM, con Ubuntu OS 10.04. La clasificación tomó aproximadamente 5:00 minutos de tiempo de cálculo, produciendo 1,741 grupos de mensajes con más de un miembro del grupo, con un

promedio de 4.13 mensajes por grupo. En comparación, un sistema de agrupación convencional que emplea un algoritmo de coincidencia de cadenas basado en hash y que se ejecuta en la misma plataforma de hardware produjo 1.617 grupos con más de un miembro en 7:07 minutos de tiempo de cálculo, con un promedio de 4.26 comentarios por grupo.

Las figs. **10-A-D** muestran los resultados de un experimento de computadora, realizado con una colección de cadenas de prueba que varían en longitud de cadena de aproximadamente 25 a 5500 caracteres. La configuración del hardware era la misma que la anterior. La Figura **10-A** muestra el tiempo que se tarda en generar histogramas de caracteres en función del número de caracteres distintos de las cadenas de prueba. La Figura **10-B** muestra el tiempo que se tarda en generar histogramas de caracteres en función de la longitud de la cadena de las cadenas de prueba. El tiempo de cálculo por histograma varió desde unos pocos microsegundos hasta aproximadamente un milisegundo, con una correlación aproximadamente lineal entre el tiempo y la longitud de la cadena.

La Figura **10-C** muestra el tiempo que se tarda para calcular un conjunto de distancias entre cadenas según una función del número de caracteres distintos de las cadenas de prueba, mientras que la Figura **10-D** muestra los mismos datos representados en función de la longitud de la cadena de las cadenas de prueba. El cálculo de las distancias entre cadenas se realizó a partir de determinaciones del recuento de caracteres, de acuerdo con las fórmulas [2-3], y varió desde unos pocos microsegundos hasta alrededor de 500 microsegundos.

La Figura **11** muestra la longitud de la cadena graficada versus un indicador de marca de tiempo, para una colección de 8676 comentarios de blog reales, que incluyen tanto spam como no spam. La Figura **12** muestra el número de caracteres distintos, graficados versus un indicador de marca de tiempo, para otra colección de 5351 comentarios reales de blog, que comprende tanto spam como no spam. Ambas Figs. **11** y **12** indican una agrupación de mensajes de spam según la marca de tiempo y la longitud de la cadena, lo que permite la identificación automática de spam como se describe anteriormente.

Quedará claro para un experto en la materia que las realizaciones anteriores pueden alterarse de muchas maneras sin apartarse del alcance de la invención. Por consiguiente, el alcance de la invención debe determinarse por las siguientes reivindicaciones.

REIVINDICACIONES

1. Un sistema informático (10) que incluye un servidor de contenido (12) y un servidor anti-spam (16) que comprende al menos un procesador programado para:

en respuesta a la recepción en el servidor de contenido (12) de una cadena de destino (42) que forma parte de una comunicación electrónica, seleccionare una pluralidad de cadenas candidatas de un corpus de cadenas de referencia, en donde la selección de la pluralidad de cadenas candidatas comprende:

comparar una longitud de cadena de la cadena de destino (42) con una longitud de cadena de una cadena de referencia del corpus, y

en respuesta, seleccionar la cadena de referencia en la pluralidad de cadenas candidatas de acuerdo con el resultado de la comparación de las longitudes de las cadenas:

en respuesta a la selección de las cadenas candidatas, realizar una primera comparación entre la cadena de destino (42) y una cadena candidata de la pluralidad de cadenas candidatas, y una segunda comparación entre la cadena de destino (42) y la cadena candidata; y

determinar en el servidor anti-spam (16) si la comunicación electrónica es spam o no spam según el resultado de la primera comparación y la segunda comparación,

en donde la primera comparación comprende comparar, para cada carácter de una pluralidad de caracteres alfanuméricos distintos, un recuento de ocurrencias de cada carácter dentro de la cadena de referencia, y

en donde la segunda comparación comprende comparar una marca de tiempo de la comunicación electrónica con una marca de tiempo de otra comunicación electrónica que contiene la cadena candidata.

2. El sistema informático de la reivindicación 1, en el que el corpus de cadenas de referencia comprende una pluralidad de grupos, cada grupo incluye un conjunto de cadenas similares, en la que cada cadena candidata de la pluralidad de cadenas candidatas es representativa de un grupo distinto, y en la que el procesador está además programado, en respuesta a la realización de la primera comparación, para seleccionar un grupo de la pluralidad de grupos y para asignar la cadena de destino al grupo seleccionado.

3. El sistema informático de la reivindicación 2, en el que el al menos un procesador está además programado para determinar si la comunicación de destino es spam o no spam según una pluralidad de marcas de tiempo, cada marca de tiempo de la pluralidad de marcas de tiempo correspondientes a un miembro del grupo seleccionado .

4. El sistema informático de la reivindicación 2, en el que el al menos un procesador está además programado para:

en respuesta a la asignación de la cadena de destino al grupo seleccionado, determinar un recuento de miembros del grupo del grupo seleccionado; y

determinar si la comunicación electrónica es spam o no spam según el recuento de miembros del grupo.

5. El sistema informático de la reivindicación 2, en el que el al menos un procesador está además programado para identificar la comunicación electrónica como perteneciente a una ola de spam seleccionada de acuerdo con el grupo seleccionado.

6. El sistema informático de la reivindicación 1, en el que la selección de la pluralidad de cadenas candidatas comprende además:

determinar un primer recuento de caracteres distintos de la cadena de destino y un segundo recuento de caracteres distintos de la cadena de referencia, y

cuando el primer recuento difiere del segundo recuento en una cantidad menor que un umbral predeterminado, seleccionar la cadena de referencia en la pluralidad de cadenas candidatas.

7. El sistema informático de la reivindicación 1, en el que la selección de la pluralidad de cadenas candidatas comprende:

determinar una primera puntuación de cadena de la cadena de destino como una función de:

$$\sum_i p_i w_i$$

donde p_I denota un código ASCII del i-th carácter de la cadena de destino, y w_I es una ponderación específica del carácter;

determinar una segunda puntuación de cadena de la cadena de referencia; y

cuando la puntuación de la primera cadena difiere de la puntuación de la segunda cadena en una cantidad menor que un umbral predeterminado, seleccionar la cadena de referencia en la pluralidad de cadenas candidatas.

8. El sistema informático de la reivindicación 1, en el que la realización de la primera comparación comprende determinar una distancia entre cadenas en función de:

$$\sum_{i \in T \cap C} w_i \left| N^i{}_T - N^i{}_C \right| ,$$

donde *T* denota el conjunto de caracteres de la cadena de destino, *C*denota el conjunto de caracteres de la cadena candidata, $N^i{}_T$ denota un recuento de ocurrencias del carácter *i* dentro de la cadena de destino, $N^i{}_C$ denota un recuento de ocurrencias del carácter *i* dentro de la cadena candidata, y en donde w_i es una ponderación específica del carácter *i.*

9. El sistema informático de la reivindicación 8, en el que la distancia entre cadenas se determina además en función de:

$$\sum_{j \in T - C} w_j \cdot c ,$$

en que el carácter *j aparece* dentro de la cadena de destino, pero no aparece dentro de la cadena candidata, w_j es una ponderación específica del carácter *j,* y *c* es un número seleccionado de acuerdo con la longitud de la cadena de la cadena de destino.

10. El sistema informático de la reivindicación 1, en el que realizar la primera comparación comprende determinar una distancia entre cadenas en función de:

$$\sum_{i \in T - C} w_i \cdot c ,$$

donde *T* denota el conjunto de caracteres de la cadena de destino, *c* denota el conjunto de caracteres de la cadena candidata, en donde carácter *i* ocurre dentro de la cadena de destino, pero no ocurre dentro de la cadena candidata, w_i es una ponderación específica del carácter *i* y *c* es un número seleccionado de acuerdo con la longitud de la cadena de la cadena de destino.

11. El sistema informático de la reivindicación 1, en el que la comunicación electrónica comprende un comentario de blog.

12. El sistema informático de la reivindicación 1, en el que la comunicación electrónica comprende un mensaje publicado en un sitio de red social.

13. Un procedimiento en un servidor de contenido (12) que comprende:

emplear al menos un procesador en el servidor de contenido (12) de un sistema informático (10) para recibir una comunicación electrónica;

en respuesta a la recepción de la comunicación electrónica, emplear al menos un procesador para extraer una cadena de destino (42) de la comunicación electrónica;

emplear el al menos un procesador para transmitir la cadena de destino (42) a un servidor anti-spam (16), y

en respuesta a la transmisión de la cadena de destino (42), emplear al menos un procesador para recibir una etiqueta de destino (50) indicativa de si la comunicación electrónica es spam o no spam, en donde la etiqueta de destino (50) se determina en el servidor anti-spam (16) y en el que la determinación de la etiqueta de destino (50) comprende:

emplear el servidor anti-spam (16) para seleccionar una pluralidad de cadenas candidatas de un corpus de cadenas de referencia, en donde la selección de la pluralidad de cadenas candidatas comprende:

comparar una longitud de cadena de la cadena de destino (42) con una longitud de cadena de un corpus de cadena de referencia, y

en respuesta, seleccionando la cadena de referencia en la pluralidad de cadenas candidatas de acuerdo con un resultado de las longitudes de comparación;

en respuesta a la selección de las cadenas candidatas, emplear el servidor anti-spam (16) para realizar una primera comparación entre la cadena de destino y una cadena candidata de la pluralidad de cadenas candidatas, y una segunda comparación entre la cadena de destino y la cadena candidata; y

emplear el servidor anti-spam (16) para determinar la etiqueta de destino (50) de acuerdo con el resultado de la primera comparación y la segunda comparación,

en donde la primera comparación comprende comparar, para cada carácter de una pluralidad de caracteres alfanuméricos distintos, un recuento de ocurrencias de cada carácter dentro de la cadena de destino con un recuento de ocurrencias de cada carácter dentro de la cadena de referencia, y

en donde la segunda comparación comprende comparar una marca de tiempo de la comunicación electrónica con una marca de tiempo de otra comunicación electrónica que contiene la cadena candidata.

14. Un procedimiento en un servidor anti-spam (16) que comprende:

en respuesta la recepción en el servidor anti-spam (16) de una cadena de destino (42) que forma parte de una comunicación electrónica de un servidor de contenido (12), emplear al menos un procesador de un sistema informático (10) para seleccionar una pluralidad de una cadena candidata de un corpus de cadenas de referencia, en la que la selección de la pluralidad de cadenas candidatas comprende:

comparar una longitud de cadena de la cadena de destino (42) con una longitud de cadena de una cadena de referencia del corpus, y

en respuesta, seleccionar la cadena de referencia en la pluralidad de cadenas candidatas

según un resultado de la comparación de longitudes de las cadenas;

en respuesta a la selección de las cadenas candidatas, empleando al menos un procesador para determinar una distancia entre cadenas que separa la cadena de destino (42) de una cadena candidata de la pluralidad de cadenas candidatas, la distancia entre cadenas se determina de acuerdo con un recuento de ocurrencias dentro de la cadena de destino (42) de cada carácter de una pluralidad de caracteres alfanuméricos distintos, y de acuerdo con el recuento de ocurrencias de cada carácter dentro de la cadena candidata; y

emplear el al menos un procesador para determinar una etiqueta de destino (50) indicativa de si la comunicación electrónica es spam o no spam según la distancia entre cadenas y según una comparación de una marca de tiempo de la comunicación electrónica con una marca de tiempo de otra comunicación electrónica que contiene la cadena candidata.

15. El procedimiento de la reivindicación 14, en el que el recuento de ocurrencias de cada carácter dentro de la cadena de destino (42) se determina sin tener en cuenta la posición de cada carácter con relación a otros caracteres dentro de la cadena de destino (42).

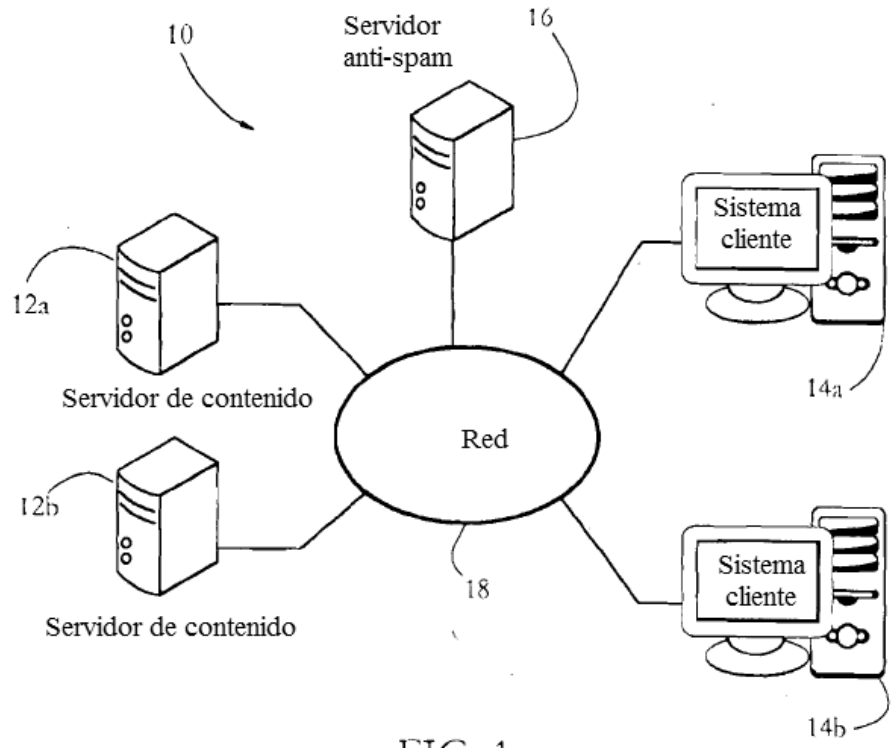
10
Servidor anti-spam
16
Sistema cliente
12a
Servidor de contenido
Red
14a
12b
18
Sistema cliente
Servidor de contenido
14b

FIG. 1

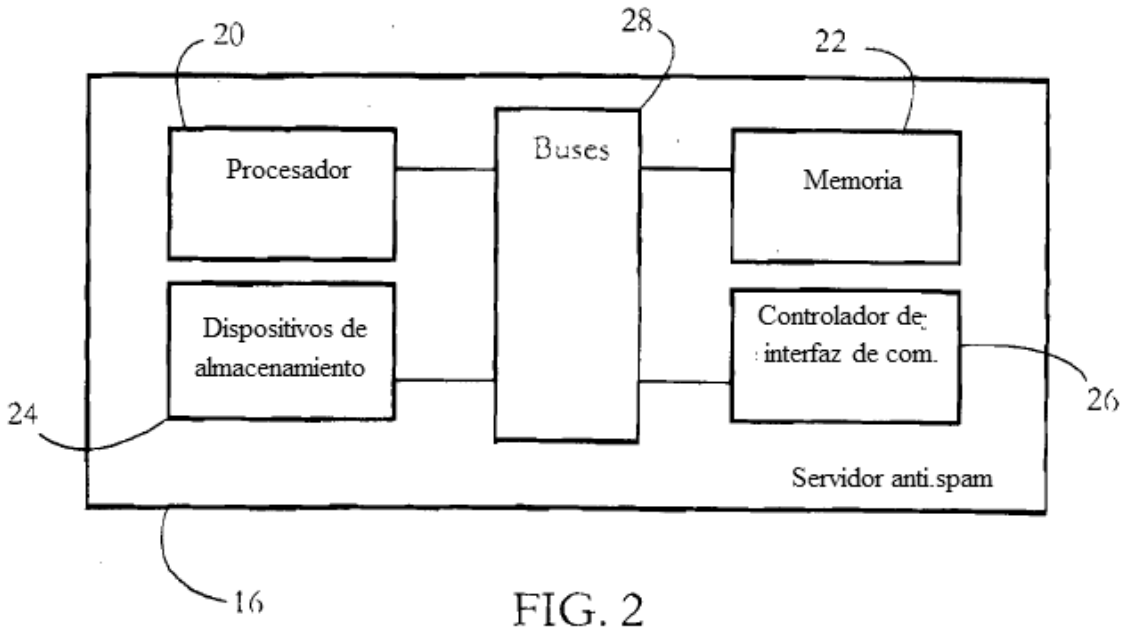
20
28
22
Procesador
Buses
Memoria
Dispositivos de almacenamiento
Controlador de interfaz de com.
24
26
Servidor anti.spam
16

FIG. 2

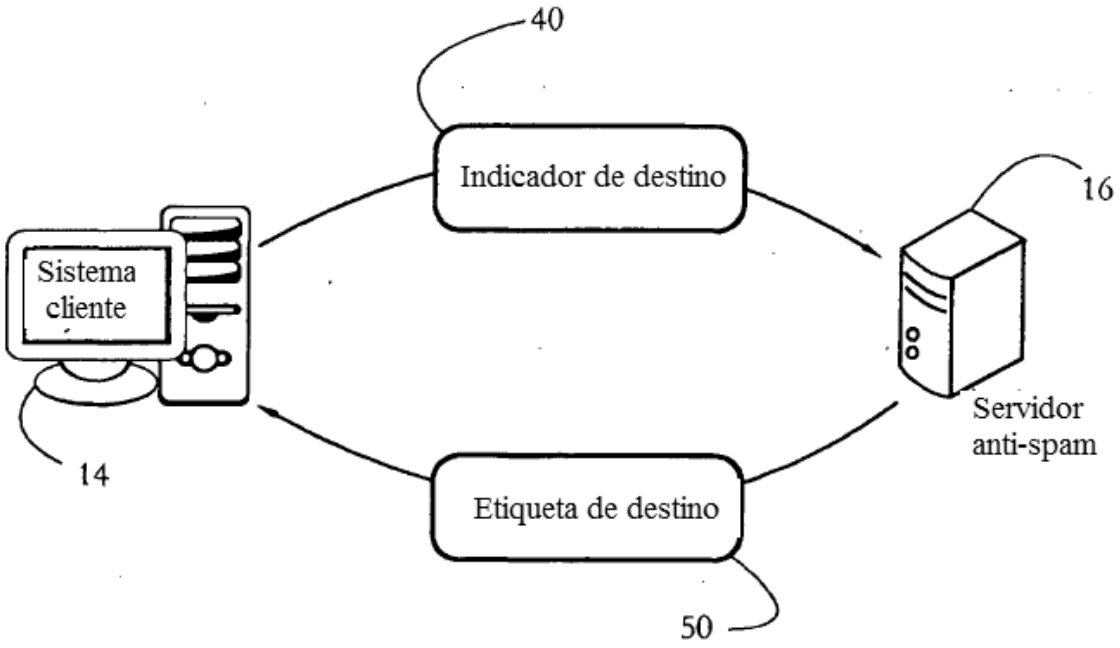
40
Indicador de destino
16
Sistema cliente
Servidor anti-spam
14
Etiqueta de destino
50

FIG. 3-A

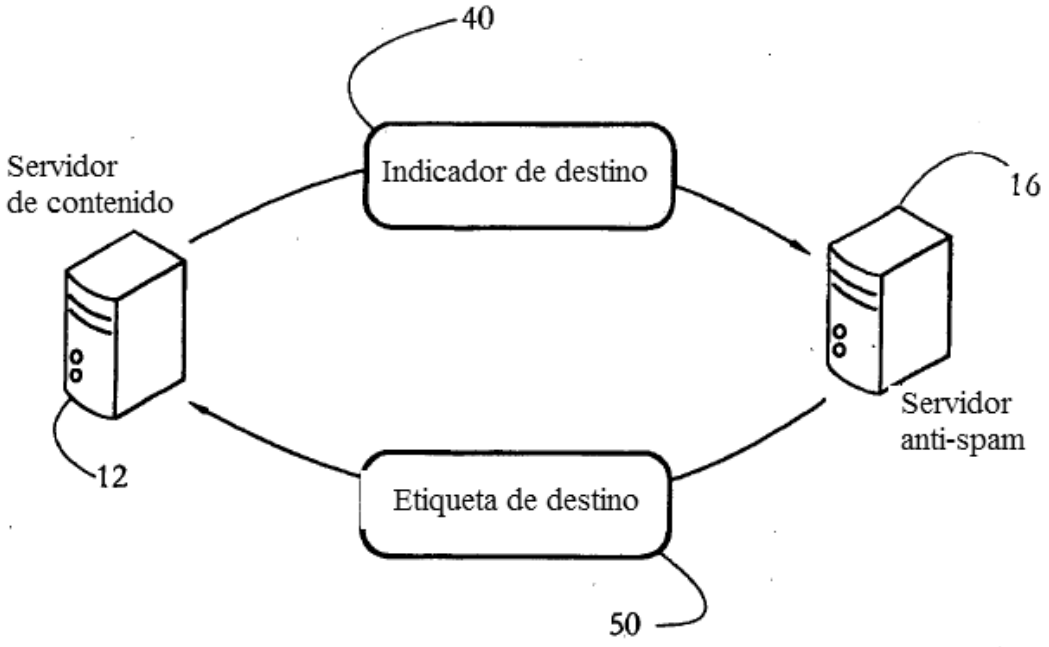
40
Servidor de contenido
Indicador de destino
16
Servidor anti-spam
12
Etiqueta de destino
50

FIG. 3-B

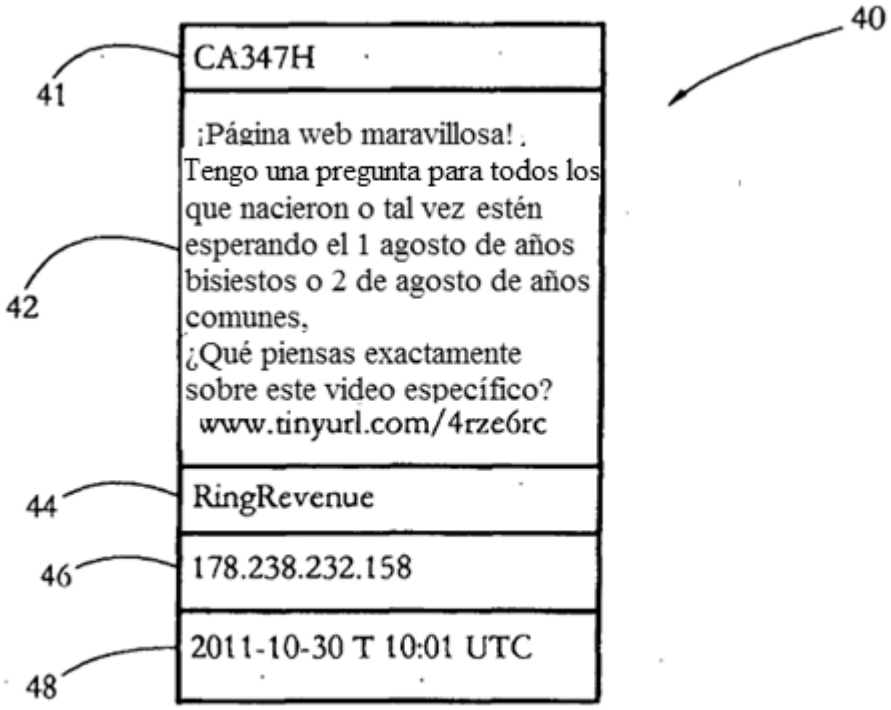
40
41
CA347H
¡Página web maravillosa!.
Tengo una pregunta para todos los que nacieron o tal vez estén esperando el 1 agosto de años bisiestos o 2 de agosto de años comunes,
¿Qué piensas exactamente sobre este video específico?
www.tinyurl.com/4rze6rc
42
44
RingRevenue
46
178.238.232.158
48
2011-10-30 T 10:01 UTC

FIG. 4

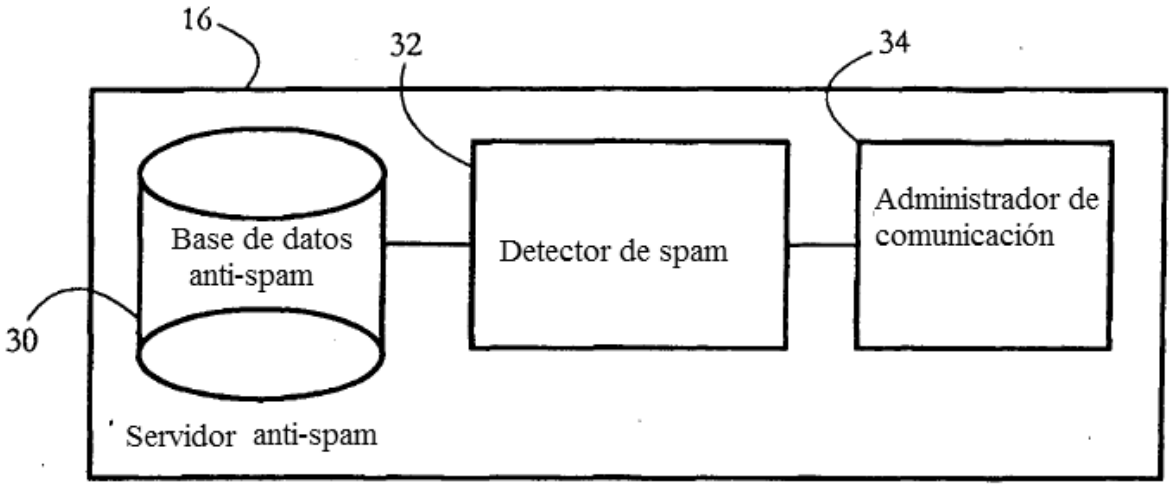
16
32
34
Base de datos anti-spam
Detector de spam
Administrador de comunicación
30
Servidor anti-spam

FIG. 5

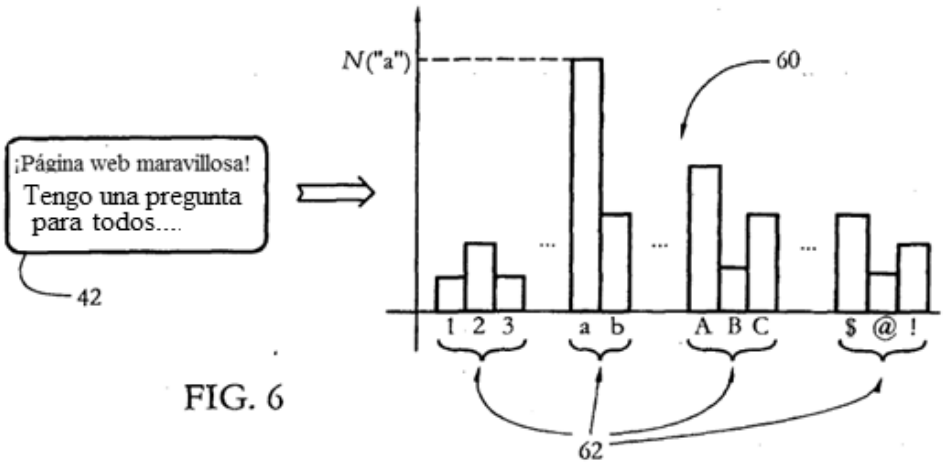
¡Página web maravillosa!
Tengo una pregunta para todos....
42
N("a")
60
1 2 3
a b
A B C
$ @ !
62

FIG. 6

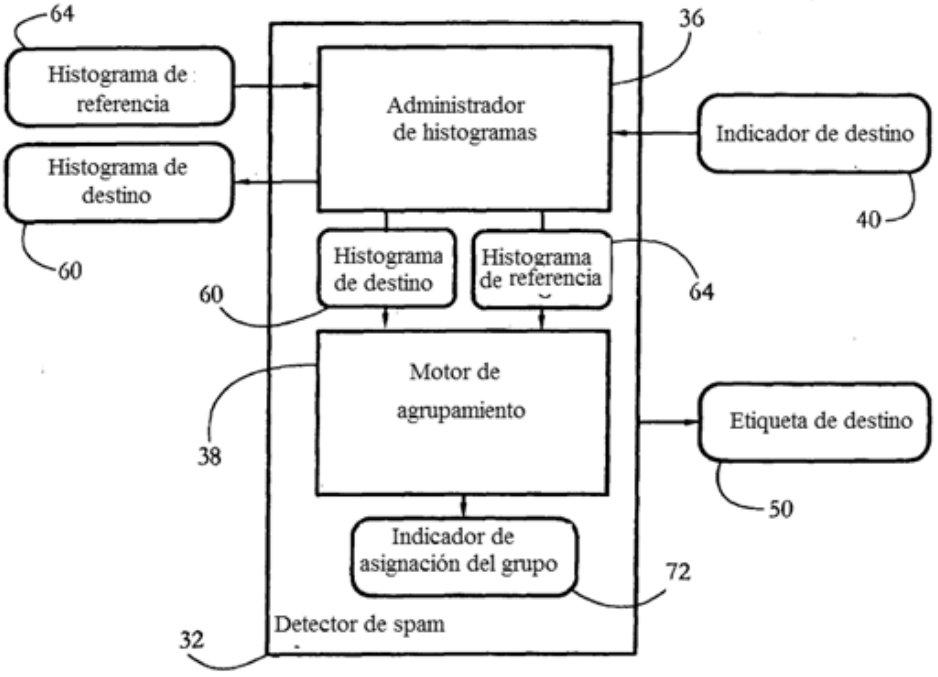
64
Histograma de referencia
Histograma de destino
60
36
Administrador de histogramas
Indicador de destino
40
Histograma de destino
Histograma de referencia
60
64
Motor de agrupamiento
38
Etiqueta de destino
50
Indicador de asignación del grupo
72
Detector de spam
32

FIG. 7

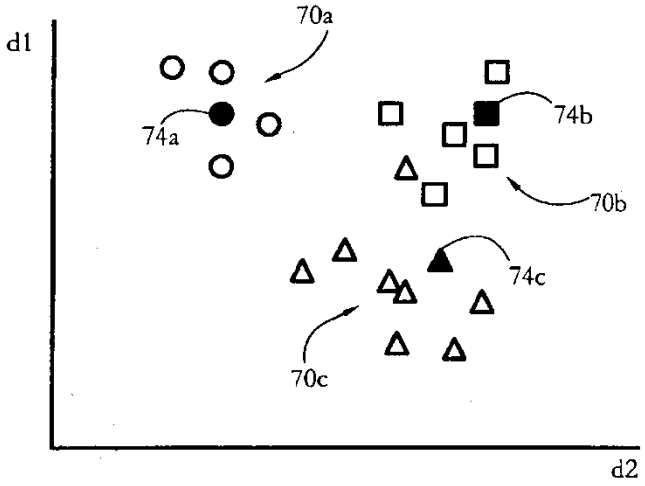

FIG. 8

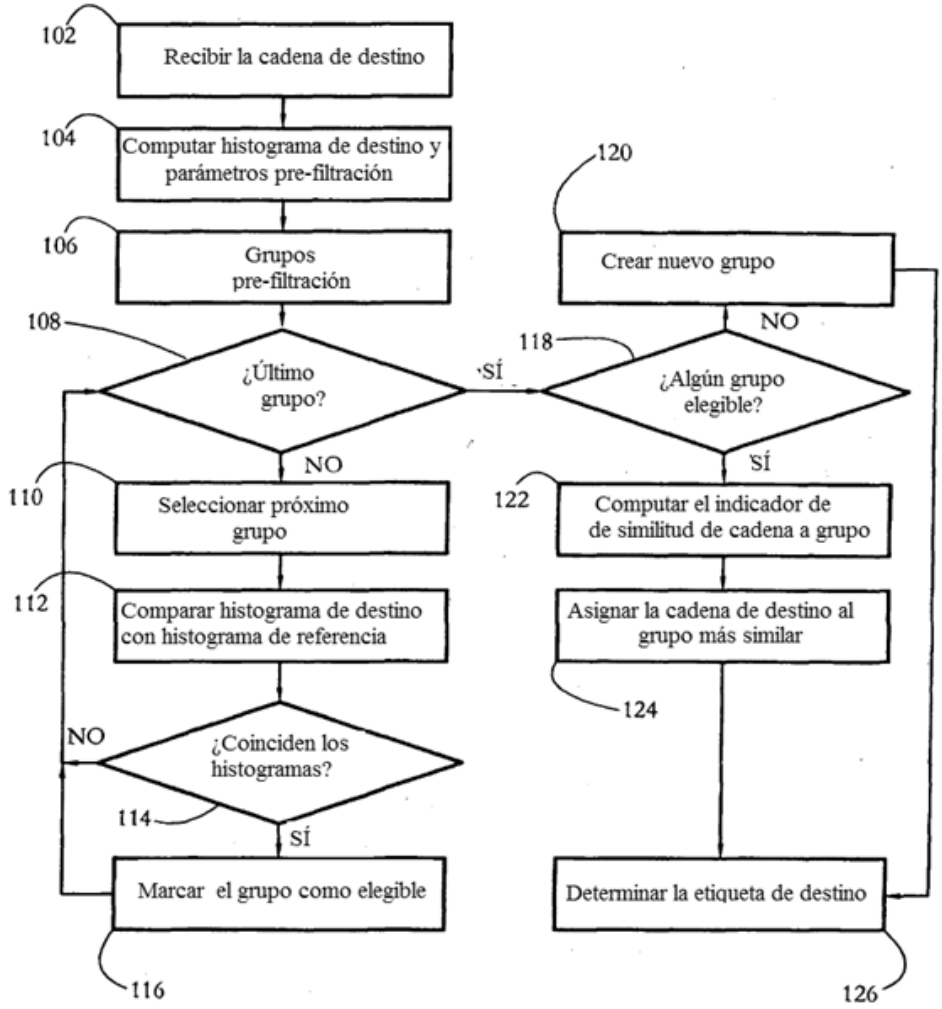
102
Recibir la cadena de destino
104
Computar histograma de destino y parámetros pre-filtración
106
Grupos pre-filtración
108
¿Último grupo?
NO
SÍ
110
Seleccionar próximo grupo
112
Comparar histograma de destino con histograma de referencia
¿Coinciden los histogramas?
NO
114
SÍ
Marcar el grupo como elegible
116
120
Crear nuevo grupo
NO
118
¿Algún grupo elegible?
SÍ
122
Computar el indicador de de similitud de cadena a grupo
Asignar la cadena de destino al grupo más similar
124
Determinar la etiqueta de destino
126

FIG. 9

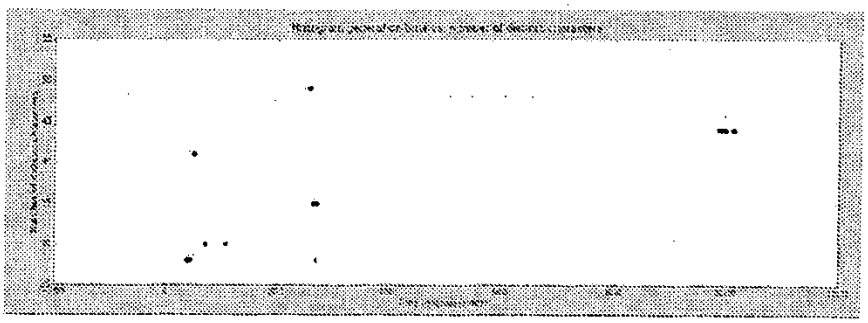

FIG. 10-A

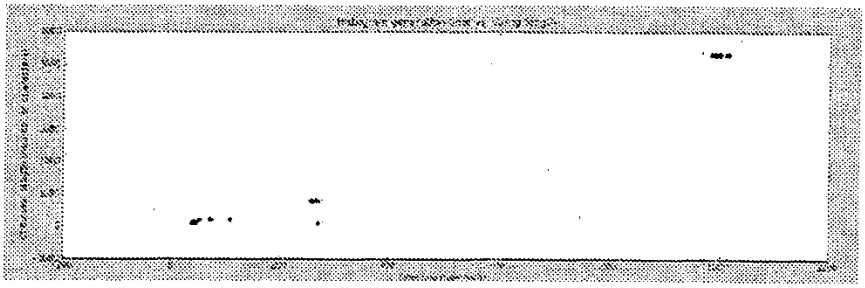

FIG. 10-B

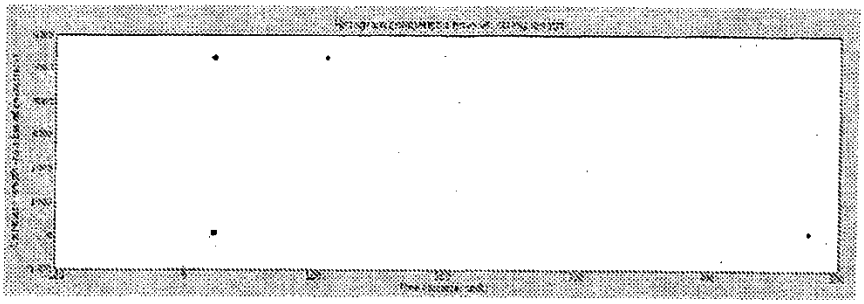

FIG. 10-C

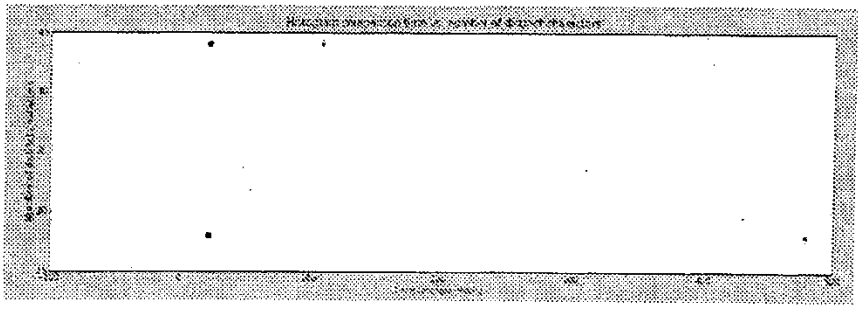

FIG. 10-D

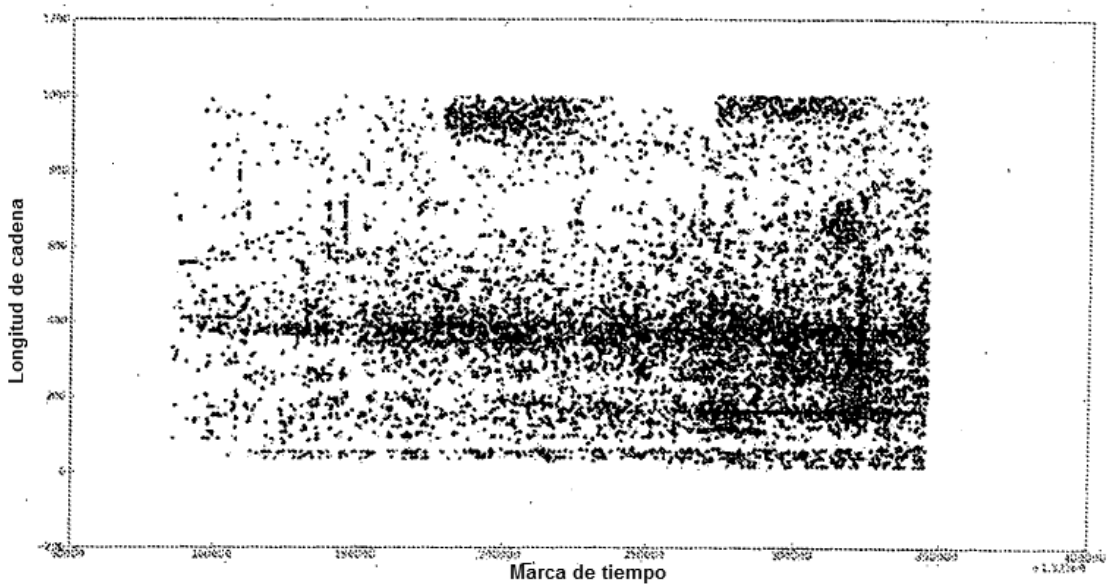
Longitud de cadena
Marca de tiempo

FIG. 11

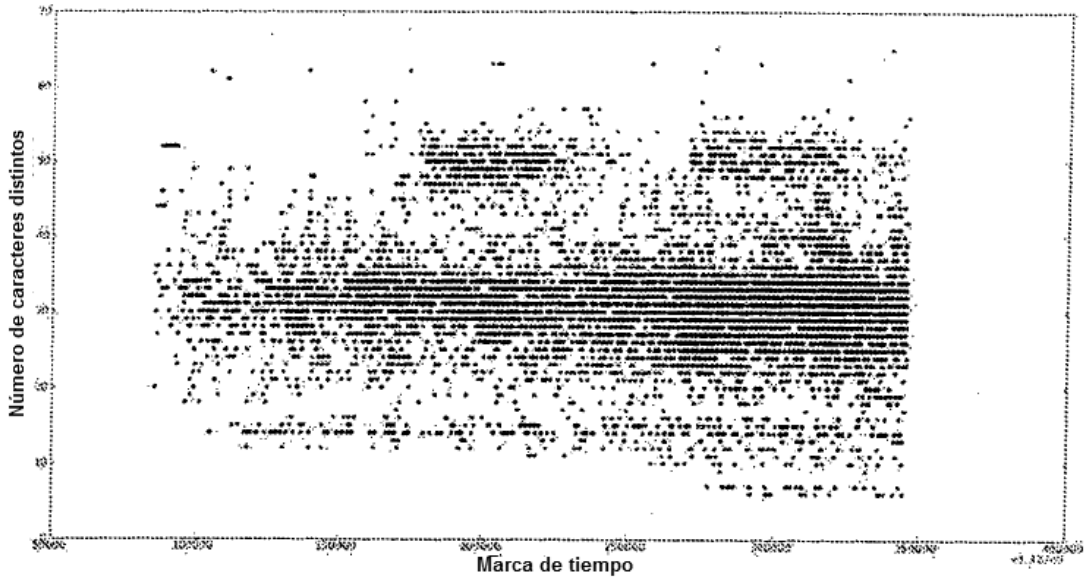
Número de caracteres distintos
Marca de tiempo

FIG. 12