(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **3 023 745**

(51) Int. Cl.:

G06F 21/62 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **20.12.2021 PCT/IB2021/062020**

(87) Fecha y número de publicación internacional: **30.06.2022 WO22137077**

(96) Fecha de presentación y número de la solicitud europea: **20.12.2021 E 21847767 (7)**

(97) Fecha y número de publicación de la concesión europea: **05.02.2025 EP 4264471**

(54) Título: **Método y sistema para redactar contenido digital no deseable**

(30) Prioridad:

21.12.2020 GB 202020296

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
03.06.2025

(73) Titular/es:

LIGHTHOUSE TECHNOLOGIES LIMITED (100.00%)
29 Craven Street
London WC2N 5NT, GB

(72) Inventor/es:

STEYNFAARDT, STEPHAN y
VAN REENEN, PIETER MEYER

(74) Agente/Representante:

DEL VALLE VALIENTE, Sonia

DESCRIPCIÓN

Método y sistema para redactar contenido digital no deseable

Campo de la invención

La invención se refiere a un método implementado por ordenador para redactar contenido digital no deseable, particularmente (pero no exclusivamente) contenido de Internet, y a un sistema para redactar tal contenido. Las realizaciones de la invención proporcionan una solución tecnológica destinada a proteger a las personas y, en particular (pero no exclusivamente), a los niños, del contenido en línea inapropiado, dañino, ofensivo u otro contenido en línea no deseable.

Antecedentes de la invención

La mayoría de las personas están expuestas a contenido digital a diario, al menos parte del cual se ofrece a través de Internet. Una forma de contenido digital es el contenido textual, visual/gráfico y/o de audio que se encuentra en línea como parte de la experiencia del usuario en un sitio web o cuando utiliza una aplicación de software basada en la web. Este tipo de contenido digital se denomina en la presente memoria "contenido de Internet" e incluye, entre otras cosas, texto, imágenes, sonidos, historias, enlaces, ventanas emergentes, vídeos y animaciones.

Internet se está convirtiendo rápidamente en una fuente fiable de información precisa y verificada. A través de Internet, ahora las personas tienen acceso a una oferta casi infinita de información y oportunidades de interacción. En la actualidad, cada vez más niños dominan la técnica de las búsquedas en Internet y, a menudo, las personas adultas no están de acuerdo sobre si deben o no permitir que los niños usen Internet.

Si bien Internet puede ser útil para ayudar a los niños a estudiar, desarrollarse y aprender la técnica de la comunicación virtual, todo lo cual se ha convertido en una parte importante de la vida, también puede exponer a los niños a contenido inapropiado para su edad. Internet, en general, no siempre se considera "seguro" para los niños y pueden existir riesgos y peligros reales para un menor de edad sin supervisión.

Los servicios en línea brindan a los niños recursos como enciclopedias, cobertura de eventos actuales y acceso a bibliotecas y otros recursos valiosos. Los niños también suelen jugar y comunicarse con sus amigos en plataformas de redes sociales como Facebook, Twitter, Snapchat, TikTok, Instagram y similares. La capacidad de "hacer clic" de un área a otra y de responder a las publicaciones de otras personas apelan a la impulsividad y la curiosidad naturales del niño, así como a sus necesidades de gratificación o retroalimentación.

A menudo, los padres enseñan a sus hijos a no hablar con extraños, a no abrir la puerta si están solos o solas en casa y a no dar información por teléfono a personas desconocidas. Los padres también supervisan con frecuencia a dónde van sus hijos, con quién juegan y a qué programas de televisión, libros o revistas están expuestos. Sin embargo, muchos padres no se dan cuenta de que se debe proporcionar el mismo nivel de orientación y supervisión para la experiencia en línea de un niño. Los padres no pueden dar por sentado que sus hijos estarán protegidos y protegidas por la supervisión o la regulación proporcionada por los servicios en línea.

Los filtros de Internet suelen incluir software que impide a los usuarios de un dispositivo informático (por ejemplo, un ordenador de sobremesa, un portátil, un teléfono móvil o una tableta) acceder a determinadas aplicaciones o sitios web, o visualizar determinadas páginas web. Se utilizan en gran medida para bloquear contenido considerado inapropiado para usuarios específicos. Este tipo de filtros se utilizan ampliamente en áreas públicas (por ejemplo, bibliotecas) y escuelas y, en cierta medida, en el lugar de trabajo. Cualquier persona que mantenga una red puede instalar filtros de Internet. Términos como filtrado, bloqueo, cierre y censura vienen a la mente cuando se describe el filtrado de Internet. Todos ellos describen que una parte de Internet o algunas aplicaciones basadas en Internet se vuelven inaccesibles debido a que el contenido se considera ofensivo, inapropiado o peligroso.

Mantener el contenido ofensivo y censurable fuera del alcance de los niños es uno de los usos más importantes de estos sistemas de filtrado. En general, los padres son conscientes de la necesidad de aplicar filtros de contenido de Internet al material que no sea apropiado para niños pequeños. Sin embargo, se apreciará que el filtrado/bloqueo de Internet no solo es aplicable a los niños, sino que también se aplica a las personas adultas. Por ejemplo, es posible que un empleador quiera bloquear o restringir el acceso a ciertos sitios web que no sean apropiados para el trabajo (por ejemplo, sitios pornográficos), que tengan un impacto negativo en el rendimiento o la eficiencia de los empleados (por ejemplo, las redes sociales) o que no sean deseables (por ejemplo, posibles estafas o actividades fraudulentas). En consecuencia, en el contexto de esta memoria descriptiva, el término "contenido digital no deseable" debería interpretarse en sentido amplio y no se limita al contenido que es inapropiado para los niños debido a su edad.

En términos de los sistemas existentes, el filtrado/bloqueo de contenido generalmente funciona mediante el uso de soluciones basadas en *hardware* o *software* para establecer reglas sobre los tipos de sitios que se pueden visitar. Utilizando palabras clave, contexto u otros puntos en común entre sitios y/o aplicaciones, el contenido se agrupa en categorías. Por ejemplo, las categorías pueden incluir deportes, juegos de azar, personas adultas, retransmisión,

compras, redes sociales y similares. En términos del funcionamiento convencional de estos sistemas de filtrado, los sitios o páginas que pertenecen a las categorías no deseadas pueden bloquearse.

Un inconveniente de los sistemas actuales es que es posible que no capten contenido digital no deseable en forma de contenido visual, por ejemplo, imágenes o vídeo o, en algunos casos, contenido de audio, especialmente cuando el contenido no deseable aparece en un sitio web, una plataforma o una aplicación marcada como "segura" en virtud de su contexto general o categoría. Por ejemplo, un niño puede tener un dispositivo con acceso a Internet en el que se aplica un filtrado para bloquear el acceso a sitios web inapropiados, tales como sitios web para personas adultas. Sin embargo, es posible que el niño siga teniendo acceso a aplicaciones de redes sociales como Facebook e Instagram, lo que lo expone a contenido visual inapropiado en forma de imágenes y vídeos, por ejemplo, imágenes o vídeos de personas adultas con ropa escasa o llamativa, sin que el sistema de filtrado identifique el contenido como potencialmente dañino (debido a que está configurado técnicamente para permitir el acceso a estas aplicaciones de redes sociales). Los sistemas existentes que el solicitante conoce se centran principalmente en el análisis del tipo de contenido (categoría) y/o texto, o en la comparación del contenido con registros almacenados, para determinar si el contenido es inapropiado, al tiempo que carecen de las capacidades técnicas necesarias para considerar el contenido visual directamente y determinar, en tiempo real, si el contenido no es deseable sin hacer referencia al tipo de contenido, el texto asociado o el contenido comparable. Similarmente, los sistemas de filtrado de contenido existentes pueden no impedir necesariamente que un usuario haga clic en un enlace incrustado no deseable o en una ventana emergente que contenga contenido digital no deseable. El documento US-2013/151346A1 es parte del estado de la técnica.

La presente invención tiene como objetivo abordar o aliviar los problemas descritos anteriormente, al menos hasta cierto punto, proporcionando una solución técnica mejorada a los problemas técnicos asociados con los sistemas existentes.

Resumen de la invención

Según la invención, en términos generales, se proporciona un método implementado por ordenador para redactar contenido digital no deseado, cuyo método comprende:

recibir, en un servidor de identificación, una solicitud de contenido que se origina en un dispositivo de usuario;

identificar una cuenta de usuario asociada a la solicitud de contenido;

etiquetar la solicitud de contenido con un identificador de restricción que sea indicativo de un parámetro de restricción asociado a un usuario del dispositivo de usuario, estando el usuario o dispositivo de usuario vinculado a la cuenta de usuario;

analizar, mediante un motor de inspección acoplado comunicativamente al servidor de identificación, el contenido digital solicitado por medio de la solicitud de contenido antes de que el contenido digital se transmita al dispositivo de usuario, en donde el contenido digital incluye contenido visual en forma de contenido de imagen y/o vídeo;

utilizar un resultado del análisis realizado mediante el motor de inspección para determinar si el contenido digital, o parte del mismo, no es deseable en función del parámetro de restricción; y

si el contenido digital o parte del mismo se clasifica como no deseable, redactar el contenido digital o parte del mismo y hacer que una versión redactada o censurada del contenido digital se transmita al dispositivo de usuario, o

si el contenido digital o parte del mismo no se clasifica como no deseable, permitir que el contenido digital se transmita al dispositivo de usuario sustancialmente sin cambios.

El parámetro de restricción puede ser un nivel de restricción. El nivel de restricción puede ser, por ejemplo, una restricción relacionada con la edad aplicada como configuración por parte de un padre/tutor al dispositivo de un niño. Esto se describe con más detalle a continuación.

Más específicamente, según un primer aspecto de la invención, se proporciona un método implementado por ordenador para redactar contenido digital no deseado, cuyo método comprende:

recibir, en un servidor de identificación, una solicitud de contenido que se origina en un dispositivo de usuario;

identificar una cuenta de usuario asociada a la solicitud de contenido;

etiquetar la solicitud de contenido con un identificador de restricción que sea indicativo de un nivel de restricción asociado a un usuario del dispositivo de usuario, estando el usuario o dispositivo de usuario vinculado a la cuenta de usuario;

analizar, mediante un motor de inspección acoplado comunicativamente al servidor de identificación, el contenido digital solicitado por medio de la solicitud de contenido antes de que el contenido digital se transmita al dispositivo de usuario, en donde el contenido digital incluye contenido visual en forma de contenido de imagen y/o vídeo, en donde el motor de inspección implementa un modelo de inspección por inteligencia artificial, y en donde el motor

de inspección se entrena, utilizando aprendizaje automático, para inspeccionar el contenido visual para determinar si el contenido visual, o parte del mismo, es apropiado para el usuario en función del nivel de restricción;

utilizar un resultado del análisis realizado mediante el motor de inspección para determinar si el contenido visual, o parte del mismo, no es deseable en función del nivel de restricción, en donde el resultado del análisis incluye una calificación asociada al contenido visual y un nivel de confianza generado mediante el motor de inspección, y en donde la calificación y el nivel de confianza se tienen en cuenta para determinar si se debe redactar el contenido visual o parte del mismo;

si el contenido visual o parte del mismo se clasifica como no deseable, redactar el contenido visual o parte del mismo y hacer que una versión redactada o censurada del contenido digital se transmita al dispositivo de usuario, en donde redactar el contenido visual o parte del mismo incluye reemplazar el contenido de imagen y/o vídeo, o partes del mismo, por contenido de imagen y/o vídeo seguro en la versión redactada o censurada, o

si el contenido visual o parte del mismo no se clasifica como no deseable, permitir que el contenido digital se transmita al dispositivo de usuario sustancialmente sin cambios.

El servidor de identificación puede ser un servidor proxy. La solicitud de contenido puede ser una solicitud web, es decir, una solicitud de contenido de Internet tal como se ha definido anteriormente, siendo el contenido de Internet un contenido digital que incluye contenido visual.

El método puede incluir establecer una conexión entre el dispositivo de usuario y el servidor de identificación de tal modo que todo el tráfico web se canalice a través del servidor de identificación antes de llegar al dispositivo de usuario. La conexión se puede establecer a través de una red privada virtual (VPN, por sus siglas en inglés). El método puede incluir, en respuesta a la determinación de que la conexión se ha interrumpido o ya no está activa, transmitir una notificación de alerta a un segundo dispositivo. El servidor puede transmitir la notificación o hacer que se transmita.

El servidor de identificación puede configurarse para identificar el usuario y/o la cuenta de usuario en función de una dirección de Protocolo de Internet (IP) del dispositivo de usuario y/o en función de las credenciales de usuario, tales como un nombre de usuario y una contraseña enviados desde el dispositivo de usuario. El servidor de identificación puede tener acceso a una o más bases de datos en las que se almacenan los detalles de la cuenta de usuario.

El servidor de identificación puede configurarse para recuperar el nivel de restricción para el usuario de la cuenta/dispositivo de usuario de una base de datos. El nivel de restricción puede ser una edad de identificación del usuario. El método puede incluir establecer la edad de identificación del usuario. También se pueden emplear otros parámetros de restricción.

El método puede incluir, después de identificar la cuenta de usuario, etiquetar todas las solicitudes de contenido posteriores procedentes del dispositivo de usuario con el identificador de restricción de tal modo que el parámetro de restricción y/o la cuenta de usuario solo necesiten identificarse una vez durante una sesión de comunicación. Por lo tanto, se puede utilizar una etiqueta para identificar la cuenta de usuario o el parámetro de restricción para su uso en futuras solicitudes. El identificador de restricción puede ser cualquier identificador adecuado, preferiblemente un identificador o token seguro.

El motor de inspección puede proporcionarse mediante una caja de inspección o clasificación. El motor de inspección puede ser un módulo o dispositivo computarizado que implemente al menos un modelo/algoritmo de inspección por inteligencia artificial (IA). El motor de inspección puede entrenarse, utilizando aprendizaje automático (ML, por sus siglas en inglés), para evaluar o inspeccionar el contenido digital, que puede incluir específicamente contenido visual, y/o para determinar si es apropiado para un usuario asociado a la cuenta de usuario en función del parámetro de restricción, por ejemplo, en función de la edad del usuario. El motor de inspección puede aplicar un modelo de IA adecuado al evaluar el contenido digital/visual.

El motor de inspección puede hacerse funcionar para recibir el contenido digital, por ejemplo, el contenido de Internet, incluido el contenido visual, al que un usuario desea acceder desde el dispositivo de usuario durante, en uso. Por tanto, el método puede incluir pasar todo el tráfico web para el dispositivo de usuario a través del motor de inspección antes de ponerlo a disposición del usuario en el dispositivo de usuario.

El dispositivo de usuario puede ser cualquier dispositivo de comunicación adecuado, tal como un ordenador, una tableta o un teléfono móvil.

Como se ha mencionado anteriormente, el contenido digital puede incluir específicamente contenido visual en forma de contenido de imagen y/o vídeo. El método puede incluir inspeccionar el contenido visual que forma parte del contenido digital y, si el motor de inspección determina que el contenido visual no es deseable para el usuario del dispositivo de usuario, el contenido visual o parte del mismo puede redactarse o censurarse. La redacción puede incluir el reemplazo del contenido de imagen y/o vídeo, o partes del mismo, por contenido de imagen y/o vídeo seguro en la versión redactada o censurada. Se prevé que el motor de inspección también pueda configurarse para analizar contenido digital en forma de audio y para censurar, redactar y/o bloquear el audio si el análisis revela que el audio no es deseable para la cuenta de usuario.

Por tanto, la redacción puede incluir el reemplazo del contenido considerado no deseable por contenido “seguro”. Por ejemplo, después de redactar el contenido visual en forma de imagen, o como parte del proceso de redacción, la imagen

puede reemplazarse por una imagen "segura", por ejemplo, una imagen en negro o en blanco y/o una imagen con un mensaje que indique que el contenido se ha redactado, o cualquier otra imagen que no sea no deseable o inapropiada. Similarmente, si un fotograma o fotogramas específicos del contenido de vídeo se consideran indeseables, o si un vídeo completo se considera no deseable, puede redactarse reemplazándolo por contenido "seguro", por ejemplo, reemplazando los fotogramas o partes no deseables del vídeo por una pantalla en negro o en blanco, y/o una pantalla con un mensaje que indique que el contenido se ha redactado. Por tanto, los términos "imagen segura", "vídeo seguro" y "contenido seguro" deben interpretarse en sentido amplio y pueden incluir cualquier contenido "alternativo" que no se considere no deseable o inseguro, etc., y que, por tanto, se inyecte en el contenido original para reemplazar el contenido original o partes del mismo.

El motor de inspección puede configurarse para generar el resultado del análisis.

El resultado puede incluir una calificación asociada al contenido visual o ambas, y un nivel de confianza generado mediante el motor de inspección.

La calificación puede ser una calificación otorgada al contenido visual, por ejemplo, una calificación o puntuación indicativa de su adecuación a la edad (por ejemplo, puntuación de adecuación o calificación de adecuación (por ejemplo, "Supervisión parental/PG", "13 años", "16 años" o similares), una calificación/puntuación que lo identifique como no deseable o una calificación/puntuación indicativa de contenido potencialmente dañino o inseguro en el contenido visual. El contenido visual solo se puede redactar si se cumplen los criterios de calificación predefinidos.

El nivel de confianza puede ser indicativo del grado o nivel de certeza de que el contenido visual o parte del mismo no es deseable, o puede ser no deseable, por ejemplo, un "porcentaje de certeza" asignado por el motor de inspección que indique "en qué medida de certeza" está el motor de inspección del contenido y/o si el contenido es inapropiado o en qué medida es inapropiado.

El servidor de identificación u otro componente/módulo pueden configurarse para determinar si el contenido digital/visual no es deseable en función del resultado.

El resultado del análisis, por ejemplo, la calificación (por ejemplo, la puntuación o calificación de adecuación) y/o el nivel de confianza pueden tenerse en cuenta para determinar si se debe redactar el contenido digital o parte del mismo.

Además, según la invención, en términos generales, se proporciona un sistema para redactar contenido digital no deseado, el sistema comprende:

un servidor de identificación que está configurado para recibir una solicitud de contenido que se origina en un dispositivo de usuario e identificar una cuenta de usuario asociada a la solicitud de contenido, en donde el servidor de identificación está configurado, además, para etiquetar la solicitud de contenido con un identificador de restricción que sea indicativo de un parámetro de restricción asociado a un usuario del dispositivo de usuario, estando el usuario o dispositivo de usuario vinculado a la cuenta de usuario; y

un motor de inspección acoplado comunicativamente al servidor de identificación, en donde el motor de inspección está configurado para analizar el contenido digital solicitado por medio de la solicitud de contenido antes de que el contenido digital se transmita al dispositivo de usuario, en donde el contenido digital incluye contenido visual en forma de contenido de imagen y/o vídeo, estando el motor de inspección o el servidor de identificación configurado para utilizar un resultado del análisis realizado mediante el motor de inspección para determinar si el contenido digital, o parte del mismo, no es deseable en función del parámetro de restricción, de tal modo que, si el contenido digital o parte del mismo se clasifica como no deseable, el contenido digital o parte del mismo se redacta y una versión redactada o censurada del contenido digital se transmite al dispositivo de usuario o, alternativamente, si el contenido digital o parte del mismo no se clasifica como no deseable, se permite que el contenido digital se transmita al dispositivo de usuario sustancialmente sin cambios.

Más específicamente, según un segundo aspecto de la invención, se proporciona un sistema para redactar contenido digital, el sistema comprende:

un servidor de identificación que está configurado para recibir una solicitud de contenido que se origina en un dispositivo de usuario e identificar una cuenta de usuario asociada a la solicitud de contenido, en donde el servidor de identificación está configurado, además, para etiquetar la solicitud de contenido con un identificador de restricción que sea indicativo de un nivel de restricción de un usuario del dispositivo de usuario, estando el usuario o dispositivo de usuario vinculado a la cuenta de usuario; y

un motor de inspección acoplado comunicativamente al servidor de identificación, estando el motor de inspección configurado para analizar el contenido digital solicitado por medio de la solicitud de contenido antes de que el contenido digital se transmita al dispositivo de usuario, en donde el contenido digital incluye contenido visual en forma de contenido de imagen y/o vídeo, en donde el motor de inspección implementa un modelo de inspección por inteligencia artificial, y en donde el motor de inspección se entrena, utilizando aprendizaje automático, para inspeccionar el contenido visual para determinar si el contenido visual o parte del mismo es apropiado para el usuario en función del nivel de restricción, estando el motor de inspección o el servidor de identificación configurado para utilizar un resultado del análisis realizado mediante

el motor de inspección para determinar si el contenido visual, o parte del mismo, no es deseable en función del nivel de restricción, en donde el resultado del análisis incluye una calificación asociada al contenido visual y un nivel de confianza generado mediante el motor de inspección, y en donde la calificación y el nivel de confianza se tienen en cuenta para determinar si se debe redactar el contenido visual o parte del mismo, de tal modo que, si el contenido visual o parte del mismo se clasifica como no deseable, el contenido visual o parte del mismo se redacta y se transmite una versión redactada o censurada del contenido digital al dispositivo de usuario o, alternativamente, si el contenido visual o parte del mismo no se clasifica como no deseable, se permite que el contenido digital se transmita al dispositivo de usuario sustancialmente sin cambios, en donde redactar el contenido visual o parte del mismo incluye reemplazar el contenido de imagen y/o vídeo, o partes del mismo, por contenido de imagen y/o vídeo seguro en la versión redactada o censurada.

Como se ha mencionado anteriormente, el nivel de restricción puede basarse en la edad del usuario y el contenido digital puede ser contenido de Internet, por ejemplo, contenido visual en Internet. Por tanto, el motor de inspección puede ser operable para investigar y clasificar el contenido de Internet y, específicamente, el contenido visual, en función de la edad de identificación del usuario.

En una realización de la invención, el sistema puede, por tanto, funcionar para clasificar el contenido digital, específicamente el contenido visual en el mismo, como seguro o inseguro según un nivel de restricción, tal como la edad de identificación del usuario. En tal realización, el sistema es operable para mostrar el contenido digital considerado seguro sin cambios en el dispositivo de usuario, en uso. Además, el sistema es operable para cambiar el contenido digital considerado inseguro mostrando una versión censurada o redactada (por ejemplo, “segura”) del mismo.

La edad de identificación de un usuario puede ser programada en el sistema por el usuario u otro usuario responsable del usuario. El servidor de identificación puede ser operable para identificar de forma segura la edad de identificación del usuario que solicita el contenido digital, en uso.

En una realización de la invención, el servidor de identificación es operable para etiquetar contenido digital, tal como contenido de Internet, según la edad de identificación del usuario. En una realización de este tipo, dicho etiquetado de contenido puede proporcionarse en forma de ejemplo añadiendo un identificador de restricción u otro identificador seguro al contenido de Internet.

En una realización de la invención, el sistema incluye una o más bases de datos operables para almacenar y proteger de forma segura la información personal de un usuario, en uso. La información personal puede incluir detalles de la cuenta como (aunque no de forma limitativa) el número de identidad, la dirección de correo electrónico, el número de teléfono móvil, la edad del usuario, los dispositivos vinculados a la cuenta de usuario, los detalles del usuario que administra la cuenta de usuario, etc.

Según un tercer aspecto de la invención, se proporciona un producto de programa informático para redactar contenido digital, comprendiendo el producto de programa informático al menos un medio de almacenamiento legible por ordenador que tiene instrucciones de programa incorporadas en el mismo, siendo las instrucciones de programa ejecutables por al menos un ordenador para hacer que el al menos un ordenador lleve a cabo el método sustancialmente como se ha descrito anteriormente. El medio de almacenamiento legible por ordenador puede ser un medio de almacenamiento no transitorio. La invención es tal como se define en las reivindicaciones adjuntas.

Breve descripción de los dibujos

La invención se describirá ahora con más detalle, a modo de ejemplo, con referencia a los dibujos adjuntos. En los dibujos:

la **figura 1** es un diagrama esquemático de un sistema para redactar contenido digital, según una realización de la invención;

la **figura 2** es otro diagrama esquemático del sistema de la figura 1, que ilustra partes de un flujo de proceso utilizado en un método de redacción de contenido digital, según una realización de la invención;

la **figura 3** es un diagrama de flujo que ilustra ciertas etapas y procesos empleados en un método de redacción de contenido digital, según una realización de la invención;

la **figura 4** es una captura de pantalla ilustrativa que ilustra el contenido digital antes de la redacción;

la **figura 5** es una captura de pantalla ilustrativa que ilustra el contenido digital después de la redacción; y

la **figura 6** es un diagrama en bloque de un sistema informático ilustrativo capaz de ejecutar un producto de programa informático para proporcionar funciones y/o acciones según al menos algunos aspectos de la invención.

Descripción detallada con referencia a los dibujos

La siguiente descripción de la invención se proporciona como una enseñanza habilitadora de la invención, es ilustrativa de los principios de la invención y no pretende limitar el alcance de la invención. Se entenderá que se pueden realizar

cambios en las realizaciones descritas y representadas sin dejar de obtener los resultados beneficiosos de la presente invención. Además, se entenderá que algunos beneficios de la presente invención pueden lograrse seleccionando algunas de las características de la presente invención sin utilizar otras características. En consecuencia, los expertos en la técnica reconocerán que las modificaciones y adaptaciones a la presente invención son posibles e, incluso, pueden ser deseables en ciertas circunstancias y forman parte de la presente invención.

En la figura 1, una realización de un sistema basado en la nube para redactar contenido de Internet se describe generalmente con referencia al número 100. El sistema de redacción basado en la nube de la figura 1 se denomina simplemente "sistema 100" a continuación para facilitar la referencia. En este ejemplo, el sistema 100 está configurado específicamente para clasificar el contenido como seguro o inseguro (no deseable) en función de la edad del usuario que intenta acceder al contenido. El objetivo es, por tanto, proteger a los niños de contenidos en línea inapropiados, dañinos, ofensivos u otros contenidos indeseables en línea. El nivel de restricción es, por lo tanto, la edad del usuario en esta realización. Sin embargo, se apreciará que se pueden emplear otros parámetros de restricción sin abandonar el ámbito de la invención, por ejemplo, comprobar si el contenido es apropiado o inapropiado para un usuario que ha iniciado sesión en un ordenador de trabajo (a menudo denominado "seguro para el trabajo" o "no seguro para el trabajo").

Además, en esta realización, el sistema 100 está configurado específicamente para analizar contenido digital en forma de contenido visual, incluyendo imágenes y/o contenido de vídeo.

En términos generales, el sistema 100 se utiliza para monitorizar el contenido visual en tiempo real, para calificar/puntuar y clasificar el contenido como permisible o no deseable, y para redactar/bloquear el contenido no deseable utilizando técnicas personalizadas de aprendizaje automático e inteligencia artificial. El sistema 100 puede inyectar/reemplazar el contenido visual con "contenido alternativo" en función de unos criterios de calificación.

En esta realización, el sistema 100 está basado en la nube y proporciona un servicio de monitorización y censura del contenido de Internet. Sin embargo, se pueden emplear realizaciones alternativas, tales como las realizaciones en las que algunos de los componentes del sistema 100 se ejecutan, por ejemplo, en instancias *Docker* (contenedores) en la infraestructura de un cliente que utiliza el servicio.

El sistema 100 incluye un servidor proxy denominado servidor 102 de identificación y también incluye una *AI-Box* (caja de IA) denominada motor 104 de inspección que está acoplada comunicativamente al servidor 102. Múltiples usuarios 110, 112, 114 pueden conectarse al sistema 100, típicamente a través de una conexión VPN, utilizando dispositivos 120, 122, 124 de comunicación adecuados. Se pueden utilizar diversos tipos de dispositivos, como se muestra en la figura 1. En uso, se puede instalar una aplicación de software en los dispositivos 120, 122, 124 y se puede crear una cuenta de usuario para cada usuario 110, 112, 114. Alternativamente, un padre puede crear una cuenta de usuario y vincular esa cuenta de usuario al dispositivo de un niño para administrar ese dispositivo. Por lo tanto, más de un usuario y dispositivo de usuario pueden estar asociados a una cuenta de usuario específica. Normalmente, el titular o administrador de la cuenta de usuario establece restricciones de edad para cada usuario y/o dispositivo de usuario de la cuenta de usuario. El contenido en línea se restringirá entonces según el límite de edad de esos dispositivos y/o asociado a cada usuario en particular (el límite de edad será el "nivel de restricción").

El sistema 100 permite a la persona que gestiona la cuenta de usuario establecer un parámetro de restricción en forma de un nivel de restricción asociado a la cuenta de usuario. Puede tratarse de un nivel de fidelidad, una puntuación o una calificación, por ejemplo, correlacionado con las calificaciones de los pósteres de películas para que sea más fácil para los usuarios: "A", "10", "13", "16" y similares. Cuando el dispositivo 120, 122, 124 de usuario envía una solicitud de contenido, el servidor 102 identifica la cuenta de usuario en cuestión y recupera un "parámetro de restricción" o "nivel de restricción" del usuario que utiliza la cuenta de usuario, por ejemplo, "13".

Cada vez que uno de los dispositivos 120, 122, 124 solicita contenido digital de Internet, por ejemplo, desde una página web 130, un servidor web 132, un almacenamiento 134 de datos o cualquier otro ordenador 136 o dispositivo, la solicitud de contenido se envía a través del servidor 102 de identificación de modo que el sistema 100 pueda determinar, en primer lugar, si el contenido solicitado y, específicamente, el contenido visual que forma parte del contenido solicitado, es apropiado en función de la edad del usuario que utiliza el dispositivo 120, 122, 124. Por tanto, el servidor 102 puede considerarse similar a un proxy intermediario (MITM, por sus siglas en inglés) que puede retransmitir y, si es necesario, alterar las comunicaciones enviadas entre un dispositivo de usuario y un recurso de Internet.

Dicho de otro modo, el sistema 100 puede configurarse para determinar una calificación o puntuación para el contenido (o partes del mismo, tales como las partes visuales), por ejemplo, la calificación de adecuación a la edad o la puntuación de adecuación a la edad del contenido y, a continuación, comparar la puntuación/calificación del contenido con el nivel de restricción del usuario en cuestión y solo permitir al usuario visualizar o acceder al contenido (o partes del mismo) si la puntuación/calificación no supera el nivel de restricción.

Para analizar el contenido digital, el motor 104 de inspección se entrena utilizando aprendizaje automático. El motor 104 puede entrenarse automáticamente para clasificar diversos tipos de contenido, tales como texto, imágenes, vídeos, historias, datos estructurados y no estructurados, y los datos apropiados pueden utilizarse para construir todos los modelos de IA que se ejecutan en el motor 104. En esta realización, y como se ha mencionado

anteriormente, el motor 104 está específicamente entrenado para analizar directamente el contenido visual independientemente del contexto o la categoría/tipo de contenido de Internet, para garantizar de este modo que el contenido visual no deseable se redacte aunque la categoría/tipo de contenido más amplia asociada al contenido de Internet (por ejemplo, su tipo de texto o página web) no sea insegura o inapropiada en sí misma.

En este ejemplo, para clasificar el contenido visual como no deseable, el modelo se creó para permitir múltiples tonos de piel y condiciones de poca luz. El modelo fue entrenado en múltiples conjuntos de imágenes:

a) SFW (seguro para el trabajo, por sus siglas en inglés): imágenes que se pueden visualizar de forma segura

b) NSFW (no es seguro para el trabajo, por sus siglas en inglés): imágenes que contienen desnudez

c) PG (supervisión parental, por sus siglas en inglés): imágenes que contienen personas en traje de baño

d) Gore (violencia gráfica): imágenes que contienen violencia gráfica

El contenido ofensivo o no deseable al que el sistema 100 está expuesto a diario a través de múltiples dispositivos puede servir como material de aprendizaje que permita al motor 104 mejorar su capacidad para clasificar y restringir correctamente el contenido a lo largo del tiempo.

En este ejemplo, el motor 104 analiza no solo el contexto del texto y la página web, sino también el contenido visual y, a continuación, genera un resultado. El resultado incluye una indicación de lo que contiene el contenido visual (por ejemplo, una imagen que sea NSFW o PG; dicho de otro modo, una calificación o puntuación para el contenido visual) junto con un nivel de confianza. El sistema 100 está configurado, además, para redactar/modificar el contenido visual que no sea apropiado para un usuario específico y transmitir una versión redactada/censurada del contenido a ese usuario.

Volviendo ahora a las figuras 2 y 3, se describe un procedimiento ilustrativo según la invención. En este ejemplo, se hace referencia al usuario 110 y el usuario 112. El usuario 110 es un niño de 14 años que utiliza el dispositivo 120 para acceder a Instagram. La usuaria 112 es una niña de 11 años que utiliza el dispositivo 122 también para acceder a Instagram.

Ambos usuarios 110, 112 están registrados en el sistema 100 como se ha descrito anteriormente y conectados al sistema 100 a través de una conexión VPN. Cada usuario 110, 112 y/o su dispositivo 120, 122 está vinculado a una cuenta de usuario con el sistema 100. Se puede utilizar cualquier conexión adecuada (por ejemplo, VPN) y reglas de firewall para canalizar todo el tráfico web (solicitudes de contenido) a través del servidor 102.

El sistema 100 puede configurarse de tal modo que se notifique a un segundo dispositivo (por ejemplo, el dispositivo de un padre) si la conexión VPN se interrumpe o se pierde. Dicho de otro modo, si en cualquier etapa del proceso se interrumpe la conexión, es posible que se notifique al dispositivo del padre por medio de un mensaje de alerta. Esto mejoró la seguridad y la eficacia del sistema 100. Esto se ilustra mediante los bloques 214 y 216 de la figura 2.

En este ejemplo, el sistema 100 ya tiene acceso a la edad de cada usuario 110, 112 o la almacena.

Para los fines de este ejemplo, se supone que ambos usuarios 110 y 112 buscan el *hashtag* (etiqueta) "#Bikini" en Instagram (véanse las figuras 4 y 5). En respuesta a la recepción de solicitudes web desde los dispositivos 120, 122 de los usuarios 110, 112, el servidor 102 identifica la cuenta de usuario asociada a cada solicitud (véanse las etapas 202 y 204 en el diagrama de bloques 200 de la figura 3).

En esta realización, el servidor 102 accedió a una base de datos para comprobar la dirección IP, el nombre de usuario y la contraseña enviados desde cada dispositivo 120, 122 y los compara con la cuenta de usuario apropiada en el sistema 100. Típicamente, al iniciar sesión en la aplicación de software que proporciona el servicio de redacción, el usuario en cuestión debe proporcionar un nombre de usuario y una contraseña, tras lo cual el servidor 102 considera tanto esto como la dirección IP para identificar al usuario y filtrar el contenido en función del nivel de "preferencias"/restricción asociado al usuario.

La solicitud web de cada dispositivo 120, 122 se etiqueta con un identificador de restricción adecuado (etapa 206 en la figura 3). En esta realización, el servidor 102 etiqueta cada solicitud con un identificador seguro que es indicativo de la edad de identificación del usuario 110, 112.

Esto permite que el motor 104 de inspección conozca la restricción de edad con la que clasificar cada solicitud de contenido. Típicamente, una vez se ha identificado un dispositivo como asociado a un usuario o una cuenta de usuario en particular durante una sesión de comunicación, el servidor 102 etiqueta todas las solicitudes web posteriores de ese dispositivo con el identificador seguro en cuestión, de modo que el parámetro de restricción (edad), el usuario o la cuenta de usuario solo necesitan identificarse una vez durante una comunicación. El identificador seguro puede ser un token que incluya datos indicativos de la edad del usuario 110, 112, lo que evita la necesidad de que el servidor 102 vuelva a comprobar la antigüedad con cada sesión de solicitud entrante (véase la flecha 218 que ilustra que la monitorización y la inspección son un proceso continuo una vez que se ha identificado al usuario). A continuación, se recupera el contenido de Internet solicitado (véase el número 150 en la figura 2) y se envía, en primer lugar, al motor 104 de inspección antes de que esté disponible para los

dispositivos 120, 122. Como se ha descrito anteriormente, el motor 104 de inspección analiza el contenido para determinar si este y, en particular, el contenido visual que forma parte del contenido, es apropiado para que los usuarios 110, 112 usen los modelos de IA que funcionan en el motor 104 (etapa 208 en la figura 3).

En este ejemplo, el motor 104 de inspección genera una puntuación o calificación con respecto a cada elemento visual del contenido digital. Véase, por ejemplo, la captura 250 de pantalla de la figura 4. En función de la puntuación o calificación otorgada a cada imagen del contenido digital, el motor 104 de inspección puede determinar que todas las imágenes son seguras para un usuario de 14 años, pero que las imágenes 252 y 254 no son seguras para los usuarios menores de 13 años debido al nivel/grado de piel expuesta, partes del cuerpo detectadas, etc. (etapa 210). En consecuencia, todo el contenido puede pasar al dispositivo 120 de usuario del usuario 110 cuando el usuario tiene 14 años, mientras que dos de las imágenes pueden redactarse antes de enviar el contenido al dispositivo 122 de usuario del usuario 112, ya que la usuaria solo tiene 11 años, como se indica mediante los números 262 y 264 en la captura 260 de pantalla de la figura 5. Véase también la figura 2, que ilustra esquemáticamente la diferencia entre el contenido visual (imágenes y vídeo) transmitido al usuario 110 (bloques 152) y al usuario 112 (bloques 154). Dicho de otro modo, los dos usuarios 110, 112 están intentando visualizar el mismo contenido web, pero la restricción de edad impuesta al usuario 112 no permite que ese usuario 112 visualice partes del contenido web, ya que esas partes se consideran indeseables y, por tanto, están redactadas.

De esta manera, el motor 104 de inspección puede recibir e inspeccionar todo el tráfico web antes de que llegue a los usuarios que utilizan el servicio, pasando contenido "seguro" a cada dispositivo de usuario en función del parámetro de restricción en cuestión mientras se redacta/censura el contenido inseguro (etapa 212 en la figura 3). Por lo tanto, el sistema 100 esencialmente "escucha" cualquier contenido entrante, particularmente contenido visual, pero también, en algunas realizaciones, otros elementos tales como texto, categorías, contexto, enlaces y ventanas emergentes, y solicita a los modelos de IA que se ejecutan en el sistema 100 que inspeccionen cada elemento. El sistema 100 determina, en función tanto del resultado de la inspección como del nivel de restricción del usuario, si el elemento puede pasarse al usuario.

El término "redactar" debería interpretarse en sentido amplio. En el ejemplo de la figura 5, las imágenes no deseables simplemente se oscurecen para formar imágenes seguras, pero se pueden emplear muchas otras técnicas. Por ejemplo, la imagen puede reemplazarse por otra imagen que sea una imagen alternativa y "segura" y/o puede incluirse un mensaje en el que se indique que el contenido se ha redactado/censurado por la seguridad del usuario.

El modelo de IA puede devolver un nivel de confianza de, por ejemplo, entre 0 y 1, lo que indica en qué medida es "cierto" que una imagen contenga algo indeseable, tales como desnudez, violencia gráfica o similares. En algunos casos, el sistema 100 puede configurarse solo para bloquear/redactar una imagen si el nivel de confianza está por encima de una determinada puntuación/umbral de confianza. Por tanto, los criterios de calificación pueden incluir, por ejemplo, redactar si el contenido visual tiene una calificación o puntuación determinada (por ejemplo, más de 13 para un niño con un nivel de restricción de "13") y si el modelo/motor arroja un nivel de confianza superior, por ejemplo, al 80 %.

Será evidente a partir de lo anterior que las realizaciones de la invención pueden proporcionar numerosas ventajas. Por ejemplo, el sistema 100 no solo considera una página web como un todo o su contexto o categoría, sino que inspecciona los elementos individuales, incluidos el vídeo y las imágenes. El sistema 100 se entrena utilizando aprendizaje automático para determinar, con un alto grado de certeza, que determinado contenido es seguro/inseguro.

Si bien los ejemplos anteriores se centran en la protección de los niños en función de su edad, las realizaciones de la invención pueden utilizarse en otras aplicaciones. Por ejemplo, el sistema 100 puede utilizarse en el lugar de trabajo para detectar contenido potencialmente fraudulento u otro contenido inseguro, y redactar el contenido en consecuencia antes de permitir que los empleados visualicen el contenido.

Las técnicas descritas anteriormente pueden implementarse en uno o más sistemas informáticos o utilizando uno de ellos, tales como el sistema informático 300 mostrado en la figura 3. El sistema informático 300 puede ser o incluir cualquier ordenador o servidor adecuado. El sistema 100 o cualquiera de los dispositivos 120, 122, 124 pueden incluir un sistema informático 300 de este tipo. El sistema informático 300 puede implementarse en entornos de computación en la nube distribuidos donde las tareas son realizadas por dispositivos de procesamiento remoto que están conectados a través de una red de comunicaciones. En un entorno de computación en la nube distribuido, los módulos de programa ejecutados por el sistema informático 300 pueden estar ubicados tanto de forma local como remota.

En el ejemplo mostrado en la figura 3, el sistema informático 300 tiene características de un ordenador de uso general. Estos componentes pueden incluir, aunque no de forma limitativa, al menos un procesador 302, una memoria 304 y un bus 306 que acopla diversos componentes del sistema 300, incluida la memoria 304, al procesador 302. El bus 306 puede tener cualquier tipo adecuado de estructura de bus. El sistema informático 300 puede incluir uno o más tipos diferentes de medios legibles, tales como medios extraíbles y no extraíbles y medios volátiles y no volátiles.

Por tanto, la memoria 304 puede incluir una memoria volátil 308 (por ejemplo, una memoria de acceso aleatorio (RAM, por sus siglas en inglés) y/o una memoria caché) y puede incluir, además, otros medios de almacenamiento, tales como un sistema 310 de almacenamiento configurado para leer y escribir en un medio no extraíble y no volátil, tal como un disco duro. Se entenderá que el sistema informático 300 también puede incluir una unidad de disco

magnético y/o una unidad de disco óptico (no mostrada), o estar acoplado a ella, para leer o escribir en medios no volátiles adecuados. Estos pueden estar conectados al bus 306 mediante una o más interfaces de medios de datos.

La memoria 304 puede configurarse para almacenar los módulos de programa 312. Los módulos 312 pueden incluir, por ejemplo, un sistema operativo, uno o más programas de aplicación, otros módulos de programa y datos de programa, cada uno de los cuales puede incluir una implementación de un entorno de red. Los componentes del sistema informático 300 pueden implementarse como módulos 312 que, generalmente, llevan a cabo funciones y/o metodologías de las realizaciones de la invención tal como se describen en la presente memoria. Se apreciará que las realizaciones de la invención pueden incluir una pluralidad de sistemas informáticos 300 o implementarse mediante ellos, que pueden estar acoplados comunicativamente entre sí.

El sistema informático 300 puede estar acoplado operativamente de forma comunicativa a al menos un dispositivo externo 314. Por ejemplo, el sistema informático 300 puede comunicarse con los dispositivos externos 314 en forma de módem, teclado y pantalla. Estas comunicaciones pueden efectuarse a través de interfaces 316 de entrada/salida (E/S) adecuadas.

El sistema informático 300 también puede configurarse para comunicarse con al menos una red 320 (por ejemplo, Internet o una red de área local) a través de un dispositivo 318 de interfaz de red/adaptador de red. El dispositivo de interfaz de red 318 puede comunicarse con los otros elementos del sistema informático 310, como se describió anteriormente, a través del bus 306.

Los componentes mostrados y descritos con referencia a la figura 3 son solo ejemplos y se entenderá que se pueden utilizar otros componentes como alternativas a los mostrados o junto con ellos.

Los aspectos de la presente invención pueden incorporarse como un sistema, método y/o producto de programa informático. En consecuencia, los aspectos de la presente invención pueden adoptar la forma de *hardware, software* y/o una combinación de *hardware* y *software* que, generalmente, pueden denominarse "componentes", "unidades", "módulos", "sistemas", "elementos" o similares en la presente memoria.

Además, los aspectos de la presente invención pueden adoptar la forma de un producto de programa informático incorporado en uno o más medios de almacenamiento legibles por ordenador que tengan incorporado un código de programa legible por ordenador. Un medio de almacenamiento legible por ordenador puede ser, por ejemplo, un sistema, aparato o dispositivo electrónico, magnético, óptico, electromagnético, infrarrojo o semiconductor, o cualquier combinación adecuada de los anteriores. En el contexto de esta especificación, un medio de almacenamiento legible por ordenador puede ser cualquier medio adecuado capaz de almacenar un programa para su ejecución o en conexión con un sistema, aparato o dispositivo. El código/las instrucciones de programa pueden ejecutarse en un único dispositivo, en una pluralidad de dispositivos (por ejemplo, en dispositivos locales y remotos), como un único programa o como parte de un sistema/paquete más grande.

La presente invención puede llevarse a cabo en cualquier forma adecuada de sistema informático, incluidos un ordenador independiente, dispositivos móviles y/o procesadores que participen en una red de ordenadores. Las realizaciones/los aspectos de la presente invención pueden llevarse a cabo en un dispositivo móvil y, por tanto, el término "ordenador" debería interpretarse de manera suficientemente amplia como para incluir un dispositivo de comunicación móvil, y el término "legible por ordenador" debería interpretarse en el sentido de incluir código legible por dispositivo móvil, almacenamiento, etc. Por tanto, los sistemas informáticos programados con instrucciones que incorporan los métodos y/o sistemas descritos en la presente memoria, los sistemas informáticos programados para realizar aspectos de la presente invención y/o los medios que almacenan instrucciones legibles por ordenador para convertir un ordenador, dispositivo móvil o similar de uso general en un sistema basado en aspectos de la presente invención, puede estar dentro del alcance de la presente invención. El término "implementado por ordenador" también debería interpretarse en sentido amplio y puede incluir específicamente métodos, procesos y/o técnicas implementados por dispositivos móviles o en ellos.

Los gráficos y/o diagramas incluidos en las figuras ilustran ejemplos de implementaciones de uno o más sistemas, métodos y/o productos de programa informático según una o más realizaciones de la presente invención. Debería entenderse que uno o más bloques de las figuras pueden representar un componente, un segmento o una porción de código que comprenda una o más instrucciones ejecutables para implementar una o más funciones lógicas especificadas. En algunas implementaciones alternativas, las acciones o funciones identificadas en los bloques pueden ocurrir en un orden diferente al mostrado en las figuras o pueden ocurrir simultáneamente.

Se entenderá que los bloques o etapas mostrados en las figuras pueden implementarse mediante componentes del sistema o instrucciones de programas informáticos. Las instrucciones pueden proporcionarse a un procesador de cualquier ordenador u otro aparato adecuado de tal modo que las instrucciones, que pueden ejecutarse mediante el procesador del ordenador u otro aparato, establezcan o generen medios para implementar las funciones o acciones identificadas en las figuras.

REIVINDICACIONES

1. Un método implementado por ordenador para redactar contenido digital no deseado, que comprende:

establecer una conexión entre un dispositivo de usuario y un servidor de identificación de tal modo que todo el tráfico web se canalice a través del servidor de identificación antes de llegar al dispositivo de usuario, lo que incluye pasar todo el tráfico web para el dispositivo de usuario a través de un motor de inspección antes de ponerlo a disposición de un usuario en el dispositivo de usuario, estando el motor de inspección acoplado comunicativamente al servidor de identificación;
recibir, en el servidor de identificación, una solicitud de contenido que se origina en el dispositivo de usuario;
identificar una cuenta de usuario asociada a la solicitud de contenido;
etiquetar la solicitud de contenido con un identificador de restricción que sea indicativo de un nivel de restricción asociado al usuario del dispositivo de usuario, estando el usuario o dispositivo de usuario vinculado a la cuenta de usuario;
analizar, mediante el motor de inspección, el contenido digital solicitado por medio de la solicitud de contenido antes de que el contenido digital se transmita al dispositivo de usuario, en donde el contenido digital incluye contenido visual en forma de contenido de imagen y/o vídeo, en donde el motor de inspección implementa un modelo de inspección por inteligencia artificial, y en donde el motor de inspección se entrena, utilizando aprendizaje automático, para inspeccionar el contenido visual para determinar si el contenido visual, o parte del mismo, es apropiado para el usuario en función del nivel de restricción;
utilizar un resultado del análisis realizado mediante el motor de inspección para determinar si el contenido visual, o parte del mismo, no es deseable en función del nivel de restricción, en donde el resultado del análisis incluye una calificación asociada al contenido visual y un nivel de confianza generado mediante el motor de inspección, y en donde la calificación y el nivel de confianza se tienen en cuenta para determinar si se debe redactar el contenido visual o parte del mismo;
si el contenido visual o parte del mismo se clasifica como no deseable, redactar el contenido visual o parte del mismo y hacer que una versión redactada o censurada del contenido digital se transmita al dispositivo de usuario, en donde redactar el contenido visual o parte del mismo incluye reemplazar el contenido de imagen y/o vídeo, o partes del mismo, por contenido de imagen y/o vídeo seguro en la versión redactada o censurada, o
si el contenido visual o parte del mismo no se clasifica como no deseable, permitir que el contenido digital se transmita al dispositivo de usuario sustancialmente sin cambios.

2. El método según la reivindicación 1, en donde el servidor de identificación es un servidor proxy y la solicitud de contenido es una solicitud web.

3. El método según la reivindicación 2, que incluye, en respuesta a la determinación de que la conexión se ha interrumpido o ya no está activa, transmitir una notificación de alerta a un segundo dispositivo.

4. El método según la reivindicación 1, en donde el servidor de identificación está configurado para identificar el usuario y/o la cuenta de usuario en función de una dirección de protocolo de Internet (IP) del dispositivo de usuario y/o en función de las credenciales de usuario enviadas desde el dispositivo de usuario.

5. El método según la reivindicación 1, en donde el nivel de restricción es una edad de identificación del usuario y en donde el método incluye determinar si el contenido visual, o parte del mismo, no es deseable en función de la edad de identificación.

6. El método según la reivindicación 1, que incluye, después de identificar la cuenta de usuario, etiquetar todas las solicitudes de contenido posteriores procedentes del dispositivo de usuario con el identificador de restricción.

7. Un sistema para redactar contenido digital no deseado, que comprende:

un servidor de identificación que está configurado para establecer una conexión con un dispositivo de usuario de tal modo que todo el tráfico web se canalice a través del servidor de identificación antes de llegar al dispositivo de usuario, y en donde el servidor de identificación está configurado para recibir una solicitud de contenido que se origina en un dispositivo de usuario e identificar una cuenta de usuario asociada a la solicitud de contenido, en donde el servidor de identificación está configurado, además, para etiquetar la solicitud de contenido con un identificador de restricción que sea indicativo de un nivel de restricción de un usuario del dispositivo de usuario, estando el usuario o dispositivo de usuario vinculado a la cuenta de usuario; y

un motor de inspección acoplado comunicativamente al servidor de identificación, en donde todo el tráfico web del dispositivo de usuario pasa a través del motor de inspección antes de ponerlo a disposición del usuario en el dispositivo de usuario, estando el motor de inspección configurado para analizar el contenido digital solicitado por medio de la solicitud de contenido antes de que el contenido digital se transmita al dispositivo de usuario, en donde el contenido digital incluye contenido visual en forma de contenido de imagen y/o vídeo, en donde el motor de inspección implementa un modelo de inspección por inteligencia artificial, y en donde el motor de inspección se entrena, utilizando aprendizaje automático, para inspeccionar el contenido visual para determinar si el contenido visual o parte del mismo es apropiado para el usuario en función del nivel de restricción, estando el motor de inspección o el servidor de identificación configurado para utilizar un resultado del análisis realizado mediante el motor de inspección para determinar si el contenido visual, o parte del mismo, no es deseable en función del nivel de restricción, en donde el resultado del análisis incluye una calificación asociada al contenido visual y un nivel de confianza generado mediante el motor de inspección, y en donde la calificación y el nivel de confianza se tienen en cuenta para determinar si se debe redactar el contenido visual o parte del mismo, de tal modo que, si el contenido visual o parte del mismo se clasifica como no deseable, el contenido visual o parte del mismo se redacta y se transmite una versión redactada o censurada del contenido digital al dispositivo de usuario o, alternativamente, si el contenido visual o parte del mismo no se clasifica como no deseable, se permite que el contenido digital se transmita al dispositivo de usuario sustancialmente sin cambios, en donde redactar el contenido visual o parte del mismo incluye reemplazar el contenido de imagen y/o vídeo, o partes del mismo, por contenido de imagen y/o vídeo seguro en la versión redactada o censurada.

8. El sistema según la reivindicación 7, en donde el servidor de identificación es un servidor proxy y la solicitud de contenido es una solicitud web.

9. El sistema según la reivindicación 8, en donde el servidor de identificación está configurado para transmitir una notificación de alerta a un segundo dispositivo en respuesta a la determinación de que la conexión se ha interrumpido o ya no está activa.

10. El sistema según la reivindicación 7, en donde el servidor de identificación está configurado para identificar el usuario y/o la cuenta de usuario en función de una dirección de protocolo de Internet (IP) del dispositivo de usuario y/o en función de las credenciales de usuario enviadas desde el dispositivo de usuario.

11. El sistema según la reivindicación 7, en donde el nivel de restricción es una edad de identificación del usuario, y en donde el motor de inspección está configurado para determinar si el contenido visual, o parte del mismo, no es deseable en función de la edad de identificación.

12. El sistema según la reivindicación 7, en donde el servidor de identificación está configurado para etiquetar todas las solicitudes de contenido posteriores procedentes del dispositivo de usuario con el identificador de restricción después de identificar la cuenta de usuario.

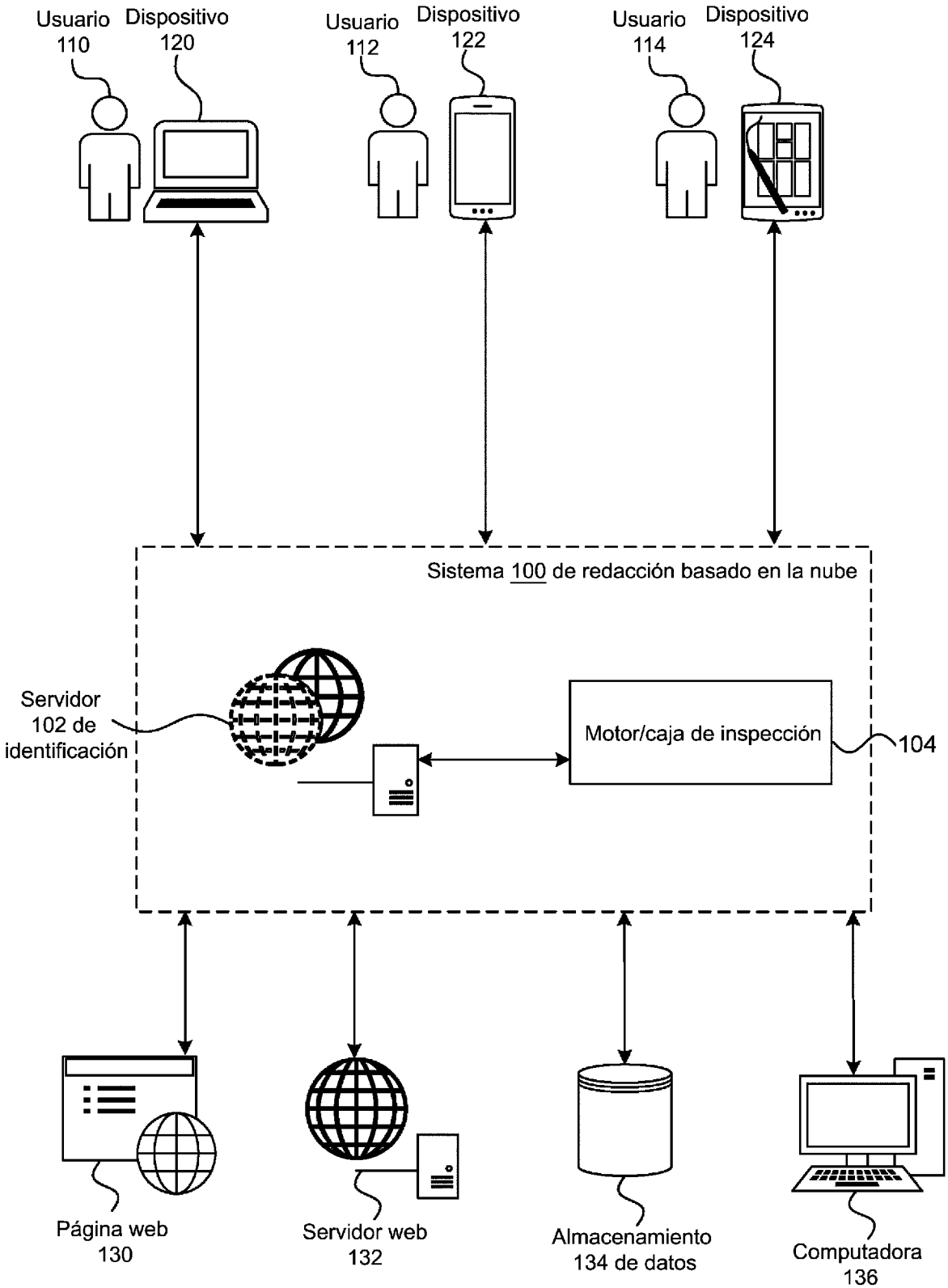
Usuario 110
Dispositivo 120
Usuario 112
Dispositivo 122
Usuario 114
Dispositivo 124
Sistema 100 de redacción basado en la nube
Servidor 102 de identificación
Motor/caja de inspección
104
Página web 130
Servidor web 132
Almacenamiento 134 de datos
Computadora 136

Figura 1

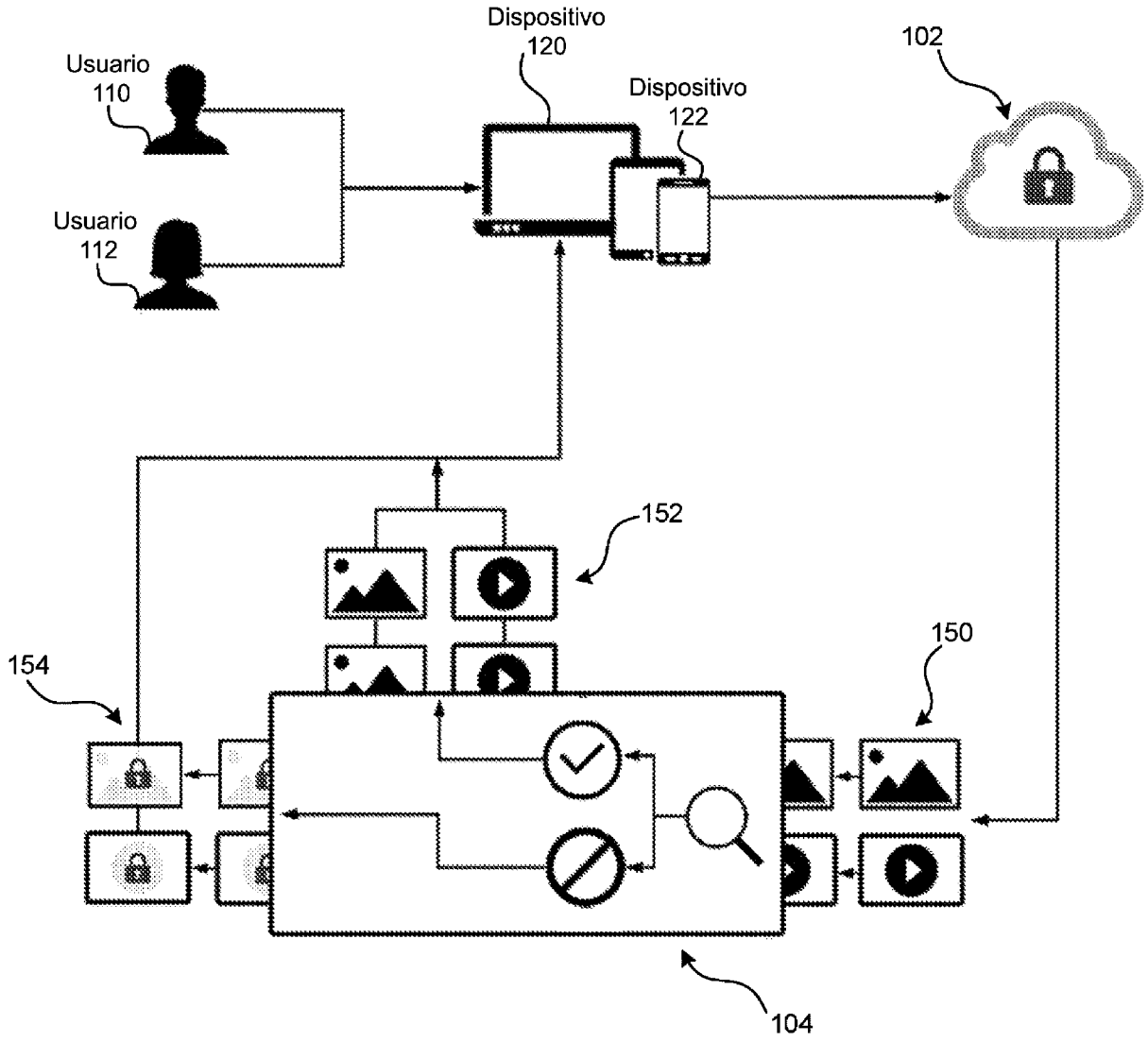
Usuario
110
Usuario
112
Dispositivo
120
Dispositivo
122
102
152
154
150
104

Figura 2

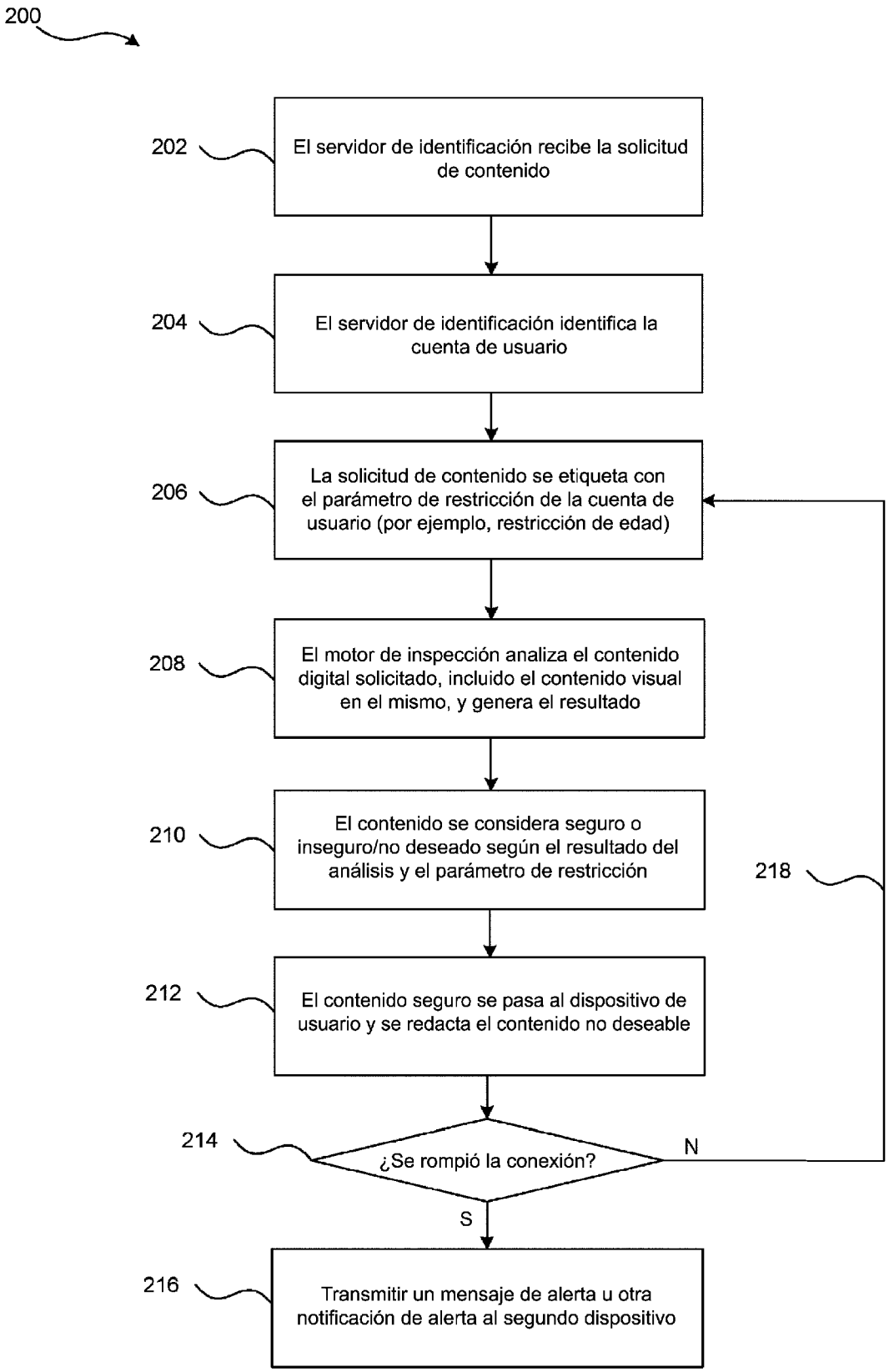
200
202
El servidor de identificación recibe la solicitud de contenido
204
El servidor de identificación identifica la cuenta de usuario
206
La solicitud de contenido se etiqueta con el parámetro de restricción de la cuenta de usuario (por ejemplo, restricción de edad)
208
El motor de inspección analiza el contenido digital solicitado, incluido el contenido visual en el mismo, y genera el resultado
210
El contenido se considera seguro o inseguro/no deseado según el resultado del análisis y el parámetro de restricción
218
212
El contenido seguro se pasa al dispositivo de usuario y se redacta el contenido no deseable
214
¿Se rompió la conexión?
N
S
216
Transmitir un mensaje de alerta u otra notificación de alerta al segundo dispositivo

Figura 3

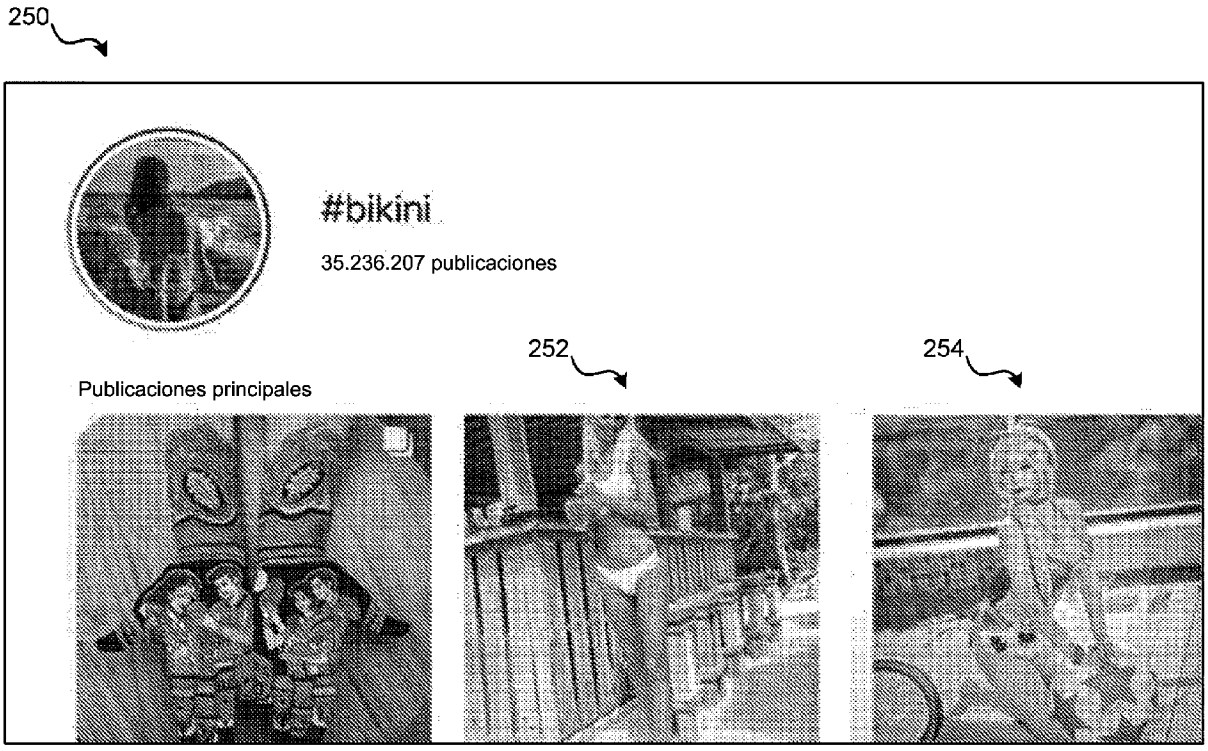

Figura 4

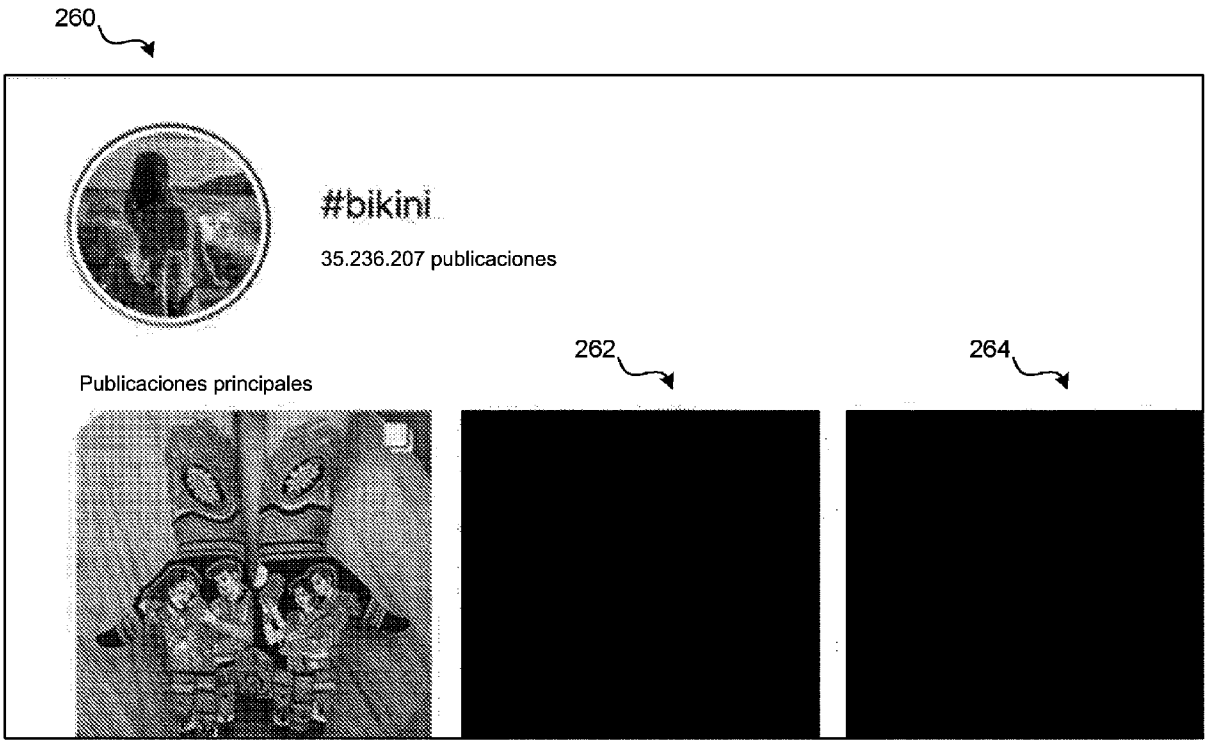

Figura 5

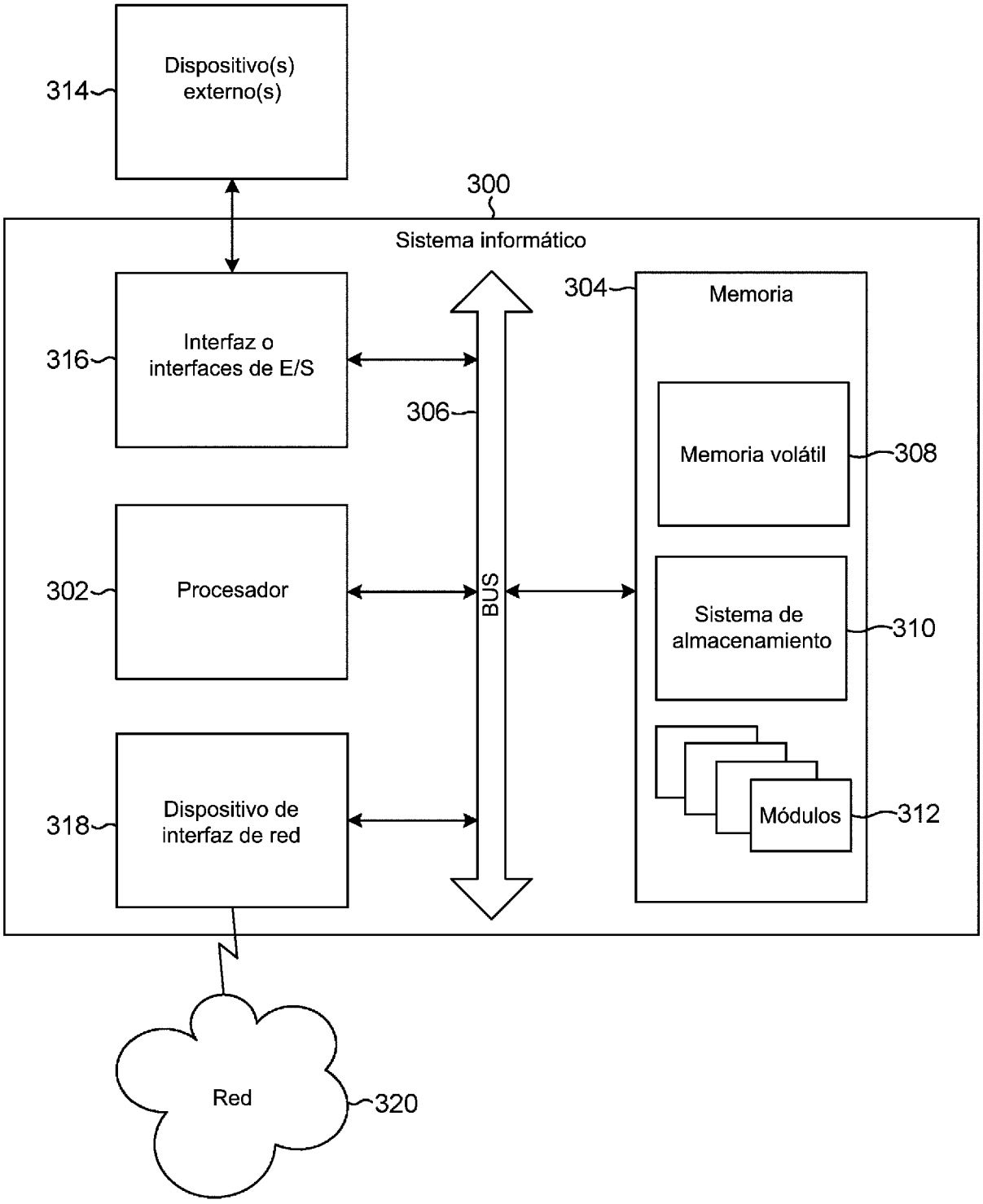
Dispositivo(s) externo(s)
314
300
Sistema informático
304
Memoria
316
Interfaz o interfaces de E/S
306
Memoria volátil
308
302
Procesador
BUS
Sistema de almacenamiento
310
318
Dispositivo de interfaz de red
Módulos
312
Red
320

Figura 6