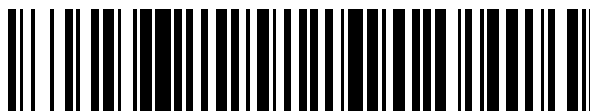


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 724 001**

51 Int. Cl.:

G06F 17/22 (2006.01)

G06K 9/00 (2006.01)

G06F 21/62 (2013.01)

G06F 17/30 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.03.2016** **E 16159615 (0)**

97 Fecha y número de publicación de la concesión europea: **20.02.2019** **EP 3173965**

54 Título: **Sistema y procedimiento de habilitación de enmascaramiento de datos para documentos web**

30 Prioridad:

26.11.2015 IN 4444MU2015

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

05.09.2019

73 Titular/es:

**TATA CONSULTANCY SERVICES LIMITED
(100.0%)**

**Nirmal Building 9th Floor Nariman Point
Mumbai 400 021, Maharashtra, IN**

72 Inventor/es:

**VIDHANI, KUMAR MANSUKHLAL;
REDDY, GANGADHARA SIRIGIREDDY;
JINDAL, ARUN KUMAR;
SHUKLA, MANISH;
BANAHATTI, VIJAYANA MAHADEO y
LODHA, SACHIN PREMSUKH**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 724 001 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Sistema y procedimiento de habilitación de enmascaramiento de datos para documentos web

Reivindicación de prioridad

5 Esta solicitud de patente reivindica prioridad sobre la Solicitud de India N.º 4444/MUM/2015, presentada el 26 de noviembre de 2015.

Campo técnico

Las realizaciones en el presente documento se refieren en general a documentos web y, más particularmente, a sistema y procedimiento de habilitación de enmascaramiento de datos para los documentos web.

Antecedentes

10 Típicamente, las organizaciones poseen grandes cantidades de datos relacionados con diversos aspectos de sus negocios, tal como empleados, socios empresariales, operaciones y gestión, almacenados en bases de datos de fuentes de contenido. Los datos almacenados en las bases de datos a menudo se usan para diferentes fines, tal como pruebas, formación, demostración e investigación de datos, y pueden accederse por personas dentro de la organización así como de fuera de la organización. Los datos almacenados en las bases de datos también son
15 accesibles como contenido web a través de documentos web o a través de otras interfaces. El contenido web representado en los documentos web, por lo tanto, contiene datos tanto sensibles como no sensibles. Por consiguiente, se necesita tener cuidado para garantizar que al menos los datos sensibles son inaccesibles para personas no autorizadas, ya sean de dentro de la organización o de fuera. Un fallo en hacer esto podría resultar en el robo de datos o divulgación innecesaria de información sensible. Por ejemplo, unos datos sensibles usados por un banco podrían incluir datos del cliente, tal como nombre, número de cuenta, número de tarjeta de crédito, número de
20 tarjeta de débito y dirección de los clientes. En muchos escenarios, no es aceptable efectuar las operaciones diarias del banco u otras organizaciones revelando la identidad de clientes a través de los datos del cliente. Sin embargo, en muchas situaciones, tal como para fines de formación y pruebas, los datos del cliente pueden tener que compartirse con otros empleados, incluso si los empleados no están autorizados a acceder a los datos. Esto puede
25 conducir a la divulgación de los datos sensibles.

Para superar los problemas anteriores, técnicas existentes realizan enmascaramiento de los datos sensibles en el documento web. Sin embargo, las técnicas existentes pueden no ser capaces de realizar enmascaramiento de los datos sensibles cuando hay un cambio en los datos sensibles y/o un cambio en una estructura del documento web.

30 El documento US 2014/101784 A1 desvela la recepción de un documento web de referencia, el resaltado automático de las etiquetas sensibles y la configuración por un usuario de una especificación de conversión para el documento web de referencia a base de etiquetas seleccionadas.

Sumario

Lo siguiente presenta un resumen simplificado de algunas realizaciones de la divulgación para proporcionar un entendimiento básico de las realizaciones. Este resumen no es una vista general extensiva de las realizaciones. No
35 pretende identificar elementos clave/críticos de las realizaciones o delinear el alcance de las realizaciones. Su único fin es presentar algunas realizaciones de una forma simplificada como un prelude a la descripción más detallada que se presenta a continuación. En vista de lo anterior, una realización en el presente documento proporciona un sistema y procedimiento de habilitación de enmascaramiento de datos para documentos web.

En un aspecto, se desvela un procedimiento de habilitación de enmascaramiento de datos para documentos web. En una realización, se reciben una especificación de conversión asociada con múltiples documentos web de referencia y documentos web de referencia previsualizados. La especificación de conversión es para datos asociados con etiquetas sensibles en los múltiples documentos web de referencia. Los datos asociados con las etiquetas sensibles se determinan sensibles a base de una o más reglas de política definidas por usuario. Además, se recibe un primer documento web procesado a base de la especificación de conversión, el primer documento web
45 incluye una o más etiquetas sensibles y datos asociados con las etiquetas sensibles. Adicionalmente, se determina un estado de la especificación de conversión tras el procesamiento del primer documento web.

Además, el primer documento web y los múltiples documentos web de referencia previsualizados se analizan cuando el estado de la especificación de conversión es etiqueta sensible no encontrada y/o datos asociados con una etiqueta sensible no encontrada. El primer documento web y los múltiples documentos web de referencia previsualizados se analizan para determinar un conjunto de candidatos para una etiqueta sensible y/o un conjunto de candidatos para datos asociados con una etiqueta sensible en el primer documento web. Además, la especificación de conversión se actualiza tras la determinación del conjunto de candidatos para una etiqueta sensible y/o el conjunto de candidatos para datos asociados con la etiqueta sensible en el primer documento web. También, se habilita el enmascaramiento de los datos asociados con las etiquetas sensibles en el primer documento
50 web usando la especificación de conversión actualizada.

En otro aspecto, se desvela un sistema de habilitación de enmascaramiento de datos para documentos web. En una realización, el sistema incluye uno o más procesadores y una memoria acoplada comunicativamente al procesador.

La memoria incluye un módulo de habilitación de enmascaramiento de datos. En una implementación de ejemplo, el módulo de habilitación de enmascaramiento de datos recibe una especificación de conversión asociada con múltiples documentos web de referencia y documentos web de referencia previsualizados. La especificación de conversión es para datos asociados con etiquetas sensibles en los múltiples documentos web de referencia. Los datos asociados con las etiquetas sensibles se determinan sensibles a base de una o más reglas de política definidas por usuario. Además, el módulo de habilitación de enmascaramiento de datos recibe un primer documento web procesado a base de la especificación de conversión, el primer documento web incluye una o más etiquetas sensibles y datos asociados con las etiquetas sensibles. Adicionalmente, el módulo de habilitación de enmascaramiento de datos determina un estado de la especificación de conversión tras el procesamiento del primer documento web.

Además, el módulo de habilitación de enmascaramiento de datos analiza el primer documento web y los múltiples documentos web de referencia previsualizados cuando el estado de la especificación de conversión es etiqueta sensible no encontrada y/o datos asociados con una etiqueta sensible no encontrada. El primer documento web y los múltiples documentos web de referencia previsualizados se analizan para determinar un conjunto de candidatos para una etiqueta sensible y/o un conjunto de candidatos para datos asociados con una etiqueta sensible en el primer documento web. Además, el módulo de habilitación de enmascaramiento de datos actualiza la especificación de conversión tras la determinación del conjunto de candidatos para una etiqueta sensible y/o el conjunto de candidatos para datos asociados con la etiqueta sensible en el primer documento web. También, el módulo de habilitación de enmascaramiento de datos habilita el enmascaramiento de los datos asociados con las etiquetas sensibles en el primer documento web usando la especificación de conversión actualizada.

Debería apreciarse por los expertos en la materia que cualquier diagrama de bloques en el presente documento representa vistas conceptuales de sistemas ilustrativos que incorporan los principios de la presente materia objeto. De manera similar, se aprecia que cualesquiera gráficos de flujo, diagramas de flujo, diagramas de transición de estado, pseudo código y similares representan diversos procedimientos que pueden representarse sustancialmente en medio legible por ordenador y ejecutarse por tanto por un dispositivo informático o procesador, se muestre o no explícitamente tal dispositivo informático o procesador.

Breve descripción de los dibujos

Las realizaciones en el presente documento se entienden mejor a partir de la siguiente descripción detallada con referencia a los dibujos, en los que:

- la Figura 1 ilustra un sistema de enmascaramiento de datos de documentos web, de acuerdo con una realización de la presente divulgación;
- la Figura 2 ilustra un diagrama de bloques detallado de un sistema de habilitación de enmascaramiento de datos de la Figura 1, de acuerdo con una realización de la presente divulgación;
- la Figura 3 ilustra una porción de un documento web de lenguaje de marcas de hipertexto (HTML), de acuerdo con una realización de la presente divulgación;
- la Figura 4 ilustra un código fuente de HTML de una porción subrayada en el documento web de HTML de la Figura 3, de acuerdo con una realización de la presente divulgación;
- la Figura 5 ilustra una especificación de conversión de una etiqueta sensible "Número de cuenta" de la Figura 3, de acuerdo con una realización de la presente divulgación;
- las Figuras 6A-6C ilustran diversos tipos de consultas de contexto, de acuerdo con una realización de la presente divulgación;
- la Figura 7 ilustra un documento web de HTML en forma de árbol, de acuerdo con una realización de la presente divulgación;
- la Figura 8 ilustra un documento web de HTML modificado en forma de árbol, de acuerdo con una realización de la presente divulgación;
- la Figura 9 ilustra un código fuente de HTML del documento web de HTML modificado, de acuerdo con una realización de la presente divulgación;
- la Figura 10 ilustra una porción del código fuente de HTML de la Figura 9, de acuerdo con una realización de la presente divulgación;
- la Figura 11 ilustra una porción modificada del código fuente de HTML de la Figura 10, de acuerdo con una realización de la presente divulgación; y
- la Figura 12 es un diagrama de flujo que ilustra un procedimiento de habilitación de enmascaramiento de datos para documentos web, de acuerdo con una realización de la presente divulgación.

Descripción detallada

Las realizaciones en el presente documento y las diversas características y detalles ventajosos de las mismas se explican con más detalle con referencia a las realizaciones no limitantes que se ilustran en los dibujos adjuntos y detallan en la siguiente descripción. Los ejemplos usados en el presente documento se conciben meramente para facilitar un entendimiento de formas en las que pueden practicarse las realizaciones en el presente documento y para habilitar adicionalmente que los expertos en la materia practiquen las realizaciones en el presente documento. Por consiguiente, los ejemplos no deberían interpretarse como que limitan el alcance de las realizaciones en el presente documento.

Las expresiones “datos sensibles” y “datos asociados con etiquetas sensibles” se usan de manera intercambiable a lo largo de todo el documento. Los términos “actualización” y “cambio” se usan de manera intercambiable a lo largo de todo el documento.

La Figura 1 ilustra un sistema 100 para el enmascaramiento de datos de documentos web, de acuerdo con una realización de la presente divulgación. Como se muestra en la Figura 1, el sistema 100 incluye un motor 102 de creación de especificación de conversión, un servidor 104 de aplicación y un sistema 106 de habilitación de enmascaramiento de datos. Además, el servidor 104 de aplicación incluye un motor 108 de privacidad. Adicionalmente, el sistema 106 de habilitación de enmascaramiento de datos incluye un módulo 110 de habilitación de enmascaramiento de datos.

En la operación, el motor 102 de creación de especificación de conversión crea una especificación de conversión para un documento web de referencia. La especificación de conversión es para datos asociados con etiquetas sensibles en el documento web de referencia y los datos asociados con las etiquetas sensibles se determinan sensibles a base de una o más reglas de políticas definidas por usuario. Por ejemplo, las etiquetas sensibles pueden representar un texto, una imagen, un botón funcional, un enlace de navegación, un área de texto, un cuadro de selección y similares. La especificación de conversión incluye información asociada con nodos de etiquetas sensibles en un documento web, información de vecino asociada con nodos de etiquetas sensibles, información de vecino de datos asociados con etiquetas sensibles e información asociada con una estructura del documento web. Además, el motor 102 de creación de especificación de conversión aplica la especificación de conversión creada en el documento web de referencia para obtener documentos web previsualizados.

Además, el motor 108 de privacidad desplegado en el servidor 104 de aplicación o en una red como intermediario identifica y enmascara datos sensibles en una respuesta de aplicación (es decir, un documento web) usando la especificación de conversión creada por el motor 102 de creación de especificación de conversión. Básicamente, el motor 108 de privacidad procesa el documento web aplicando la especificación de conversión. Por ejemplo, el documento web incluye una o más etiquetas sensibles y datos asociados con las etiquetas sensibles. En un ejemplo, el motor 108 de privacidad puede aplicar la especificación de conversión en el documento web cuando existe un cambio en una estructura del documento web en la vecindad de una etiqueta sensible y/o datos asociados con la etiqueta sensible. Sin embargo, el motor 108 de privacidad no puede aplicar la especificación de conversión cuando existe un cambio en una etiqueta sensible, un cambio en datos asociados con una etiqueta sensible o un cambio en una estructura del documento web.

En tales escenarios, el sistema 106 de habilitación de enmascaramiento de datos habilita el enmascaramiento de datos para el documento web si el motor 108 de privacidad falla. En una realización de ejemplo, el módulo 110 de habilitación de enmascaramiento de datos recibe la especificación de conversión asociada con el documento web de referencia y documento web previsualizado. Además, el módulo 110 de habilitación de enmascaramiento de datos recibe el documento web procesado a base de la especificación de conversión. En algunas realizaciones, el módulo 110 de habilitación de enmascaramiento de datos analiza los múltiples documentos web de referencia previsualizados para determinar reglas de política definidas por usuario. El módulo 110 de habilitación de enmascaramiento de datos a continuación determina la existencia de una etiqueta sensible en el primer documento web usando las reglas de política definidas por usuario. El módulo 110 de habilitación de enmascaramiento de datos a continuación envía una alerta al usuario cuando la etiqueta sensible existe en el primer documento web.

Adicionalmente, el módulo 110 de habilitación de enmascaramiento de datos determina un estado de la especificación de conversión tras el procesamiento del primer documento web. Además, el módulo 110 de habilitación de enmascaramiento de datos analiza el primer documento web y los múltiples documentos web de referencia previsualizados cuando el estado de la especificación de conversión es una etiqueta sensible no encontrada y/o datos asociados con una etiqueta sensible no encontrada. Por ejemplo, el módulo 110 de habilitación de enmascaramiento de datos analiza el primer documento web y los múltiples documentos web de referencia previsualizados para determinar un conjunto de candidatos para una etiqueta sensible y/o un conjunto de candidatos para datos asociados con una etiqueta sensible en el primer documento web.

En un ejemplo, el estado de la especificación de conversión es la etiqueta sensible no encontrada cuando se actualiza la etiqueta sensible. En este ejemplo, el módulo 110 de habilitación de enmascaramiento de datos compara el documento web con el documento web de referencia previsualizado para obtener un nodo de etiqueta sensible en el documento web de referencia previsualizado. Además, el módulo 110 de habilitación de enmascaramiento de datos determina una consulta de contexto para el nodo de etiqueta sensible usando información de vecino en la especificación de conversión para el nodo de etiqueta sensible. Adicionalmente, el módulo 110 de habilitación de enmascaramiento de datos ejecuta la consulta de contexto en el documento web para determinar el conjunto de candidatos para la etiqueta sensible actualizada en el documento web. En una realización, el módulo 110 de habilitación de enmascaramiento de datos compara la consulta de contexto para el nodo de etiqueta sensible con el documento web y determina el conjunto de candidatos para la etiqueta sensible actualizada a base de la comparación.

En otro ejemplo, el estado de la especificación de conversión son los datos asociados con la etiqueta sensible no encontrada cuando se actualizan los datos asociados con la etiqueta sensible. En este ejemplo, el módulo 110 de habilitación de enmascaramiento de datos compara el documento web con el documento web de referencia previsualizado para obtener un nodo de etiqueta sensible en el documento web de referencia previsualizado.

Además, el módulo 110 de habilitación de enmascaramiento de datos determina una consulta de contexto para los datos asociados con el nodo de etiqueta sensible usando información de vecino en la especificación de conversión para el nodo de etiqueta sensible. Adicionalmente, el módulo 110 de habilitación de enmascaramiento de datos ejecuta la consulta de contexto en el documento web para determinar el conjunto de candidatos para los datos asociados actualizados con la etiqueta sensible en el documento web.

En otro ejemplo más, el estado de la especificación de conversión son los datos asociados con la etiqueta sensible no encontrada cuando se actualiza la estructura del documento web con la etiqueta sensible y datos asociados con la etiqueta sensible no cambiada o actualizada. En este ejemplo, el módulo 110 de habilitación de enmascaramiento de datos compara el documento web con el documento web de referencia previsualizado para obtener un nodo de etiqueta sensible. Además, el módulo 110 de habilitación de enmascaramiento de datos determina un conjunto de candidatos para los datos asociados con una etiqueta sensible que corresponde al nodo de etiqueta sensible aumentando un ámbito de búsqueda dentro del documento web. Adicionalmente, el módulo 110 de habilitación de enmascaramiento de datos realiza análisis de proximidad en el conjunto de candidatos para los datos asociados con la etiqueta sensible para determinar el candidato correcto del conjunto de candidatos para los datos asociados con la etiqueta sensible.

En un ejemplo, el estado de la especificación de conversión son los datos asociados con la etiqueta sensible no encontrada cuando se actualiza una estructura del documento web con la etiqueta sensible y datos asociados con la etiqueta sensible cambiada. En este ejemplo, el módulo 110 de habilitación de enmascaramiento de datos compara el documento web con el documento web de referencia previsualizado para obtener un nodo de etiqueta sensible en el documento web de referencia previsualizado. Además, el módulo 110 de habilitación de enmascaramiento de datos determina una consulta de contexto para el nodo de etiqueta sensible usando información en la especificación de conversión para el nodo de etiqueta sensible. Adicionalmente, el módulo 110 de habilitación de enmascaramiento de datos extrae información asociada con cada nodo en el documento web. Además, el módulo 110 de habilitación de enmascaramiento de datos compara la consulta de contexto de la especificación de conversión con la información extraída para determinar un conjunto de candidatos para la etiqueta sensible actualizada y/o los datos asociados actualizados con la etiqueta sensible.

Además, el módulo 110 de habilitación de enmascaramiento de datos actualiza la especificación de conversión tras la determinación del conjunto de candidatos para una etiqueta sensible y/o el conjunto de candidatos para datos asociados con la etiqueta sensible en el documento web. También, el módulo 110 de habilitación de enmascaramiento de datos habilita el enmascaramiento de los datos asociados con las etiquetas sensibles en el documento web usando la especificación de conversión actualizada. Esto se explica en más detalle con referencia a la Figura 2.

La Figura 2 ilustra un diagrama de bloques de un sistema 200 de habilitación de enmascaramiento de datos (por ejemplo, el sistema 106 de habilitación de enmascaramiento de datos de la Figura 1), de acuerdo con una realización de la presente divulgación. Como se muestra en la Figura 2, el sistema 200 incluye uno o más procesador o procesadores 202 y una memoria 204 acoplados comunicativamente entre sí. El sistema 200 también incluye la interfaz o interfaces 206. Además, la memoria 204 incluye módulos el módulo 110 de habilitación de enmascaramiento de datos y otros módulos. El módulo 110 de habilitación de enmascaramiento de datos incluye un módulo 208 de captura de respuesta y acceso a registro, un módulo 210 de análisis de escenarios fallidos, un módulo 212 de generación de consulta de contexto, un módulo 214 de detector de proximidad, un módulo 216 de transformación de respuesta y un módulo 218 de actualización de especificación de conversión. Aunque la Figura 2 muestra componentes de ejemplo del sistema 200, en otras implementaciones, el sistema 200 puede contener menos componentes, componentes adicionales, diferentes componentes o componentes dispuestos de forma diferente que los representados en la Figura 2.

El procesador o procesadores 202 y la memoria 204 pueden acoplarse comunicativamente mediante un bus de sistema. El procesador o procesadores 202 pueden incluir circuitería que implementa, entre otras, funciones de audio y lógica asociadas con la comunicación. El procesador o procesadores 202 pueden incluir, entre otras cosas, un reloj, una unidad aritmética lógica (ALU) y puertas lógicas configurados para soportar la operación del procesador o procesadores 202. El procesador o procesadores 202 pueden ser una única unidad de procesamiento o un número de unidades, incluyendo todas múltiples unidades de cálculo. El procesador o procesadores 202 pueden implementarse como uno o más microprocesadores, microordenadores, microcontroladores, procesadores de señales digitales, unidades de procesamiento central, máquinas de estado, circuiterías lógicas y/o cualesquiera dispositivos que manipulan señales a base de en instrucciones operacionales. Entre otras capacidades, el procesador o procesadores 202 se configuran para buscar y ejecutar instrucciones legibles por ordenador y datos almacenados en la memoria 204.

Las funciones de los diversos elementos mostrados en la figura, incluyendo cualquier bloque funcional etiquetado como "procesador o procesadores", pueden proporcionarse a través del uso de hardware especializado así como hardware capaz de ejecutar software en asociación con software apropiado. Cuando se proporcionan por un procesador, las funciones pueden proporcionarse por un único procesador especializado, por un único procesador compartido o por una pluralidad de procesadores individuales, algunos de los cuales pueden compartirse. Además, el uso explícito del término "procesador" no debería interpretarse para referirse exclusivamente a hardware capaz de ejecutar software, y puede incluir implícitamente, sin limitación, hardware de procesador de señales digitales (DSP),

procesador de red, circuito integrado específico de la aplicación (ASIC), campo de matriz de puertas programables (FPGA), memoria de solo lectura (ROM) para el almacenamiento de software, memoria de acceso aleatorio (RAM) y almacenamiento no volátil. También puede incluirse otro hardware, convencional y/o personalizado.

La interfaz o interfaces 206 pueden incluir una diversidad de interfaces de software y hardware, por ejemplo, interfaces para dispositivo o dispositivos periféricos, tal como un teclado, un ratón, una memoria externa y una impresora. La interfaz o interfaces 206 pueden facilitar múltiples comunicaciones dentro de una amplia variedad de tipos de redes y de protocolos, incluyendo redes por cables, por ejemplo, red de área local (LAN), cable, etc., y redes inalámbricas, tal como LAN Inalámbrica (WLAN), celular o satélite. Para el fin, la interfaz o interfaces 206 pueden incluir uno o más puertos de conexión del sistema 200 a un servidor de aplicación y un motor de creación de especificación de conversión (como se muestra en la Figura 1).

La memoria 204 pueden incluir cualquier medio legible por ordenador conocido en la técnica incluyendo, por ejemplo, memoria volátil, tal como memoria de acceso aleatorio (SRAM) estática y memoria de acceso aleatorio dinámica (DRAM) y/o memoria no volátil, tal como memoria de solo lectura (ROM), ROM programable borrrable, memorias flash, discos duros, discos ópticos y cintas magnéticas. La memoria 204 puede almacenar cualquier número de piezas de información y datos usados por el sistema 200 para implementar las funciones del sistema 200. La memoria 204 puede configurarse para almacenar información, datos, aplicaciones, instrucciones o similares para habilitar que el sistema 200 efectúe diversas funciones de acuerdo con diversas realizaciones de ejemplo. Adicionalmente o como alternativa, la memoria 204 puede configurarse para almacenar instrucciones que cuando se ejecutan el procesador o procesadores 202 provoca que el sistema 200 se comporte de una manera como se describe en diversas realizaciones. La memoria 204 incluye el módulo 208 de captura de respuesta y acceso a registro, el módulo 210 de análisis de escenarios fallidos, el módulo 212 de generación de consulta de contexto, el módulo 214 de detector de proximidad, el módulo 216 de transformación de respuesta, el módulo 218 de actualización de especificación de conversión y otros módulos. Los módulos 208 a 218 incluyen rutinas, programas, objetos, componentes, estructuras de datos, etc., que realizan tareas particulares o implementan tipos de datos abstractos particulares. Los otros módulos pueden incluir programas o instrucciones codificadas que suplementan aplicaciones y funciones del sistema 200.

En una realización, el módulo 208 de captura de respuesta y acceso a registro obtiene un documento web procesado desde un motor de privacidad en el servidor de aplicación (por ejemplo, el motor 108 de privacidad de la Figura 1) junto con un estado de una especificación de conversión. El documento web puede procesarse parcial o totalmente. Por ejemplo, el estado de la especificación de conversión pueden incluir un estado de éxito y un estado fallido. El estado de éxito significa que la especificación de conversión configurada para el documento web se aplica apropiadamente. El estado fallido significa que al menos una parte de la especificación de conversión podría no aplicarse. Para la especificación de conversión con el estado fallido, el módulo 210 de análisis de escenario fallido asigna uno de los dos estados, etiqueta sensible no encontrada y datos asociados con una etiqueta sensible no encontrada. Si se cambia el contenido de una etiqueta sensible (es decir, la etiqueta sensible se actualiza), el motor de privacidad no puede encontrar la etiqueta sensible en el documento web analizado. En ese caso, se asigna a la especificación de conversión el estado de 'etiqueta sensible no encontrada'. Si no se encuentra la etiqueta sensible, no se buscan los datos que corresponden a esa etiqueta sensible. Cuando se asigna a la especificación de conversión 'valor sensible no encontrado' cuando se actualizan los datos asociados con la etiqueta sensible o una estructura de documento web se cambia con o sin actualizar una etiqueta sensible y/o datos asociados con una etiqueta sensible. La estructura del documento web puede cambiarse de tal forma que ya no encaja en una interpretación de dirección.

Además, si la especificación de conversión falló debido a un cambio en la etiqueta sensible o datos asociados con la etiqueta sensible, el módulo 212 de generación de consulta de contexto determina etiqueta sensible nueva o actualizada o datos asociados con la etiqueta sensible nueva o actualizada. El módulo 212 de generación de consulta de contexto a continuación ejecuta la consulta de contexto con el documento web procesado (registrado por el motor de privacidad). El resultado de consulta de contexto proporciona un conjunto de candidatos para la etiqueta sensible actualizada o datos asociados con la etiqueta sensible. Por ejemplo, una porción de una página 300 de lenguaje de marcas de hipertexto (HTML) del documento web se muestra en la Figura 3 y un código 400 fuente de HTML de una porción subrayada de la Figura 3 se muestra en la Figura 4. La especificación 500 de conversión para la etiqueta sensible de "Número de cuenta:" de la Figura 3 se muestra en la Figura 5 que se genera por el motor de creación de especificación de conversión (por ejemplo, el motor 102 de creación de especificación de conversión de la Figura 1). En este ejemplo, si "Número de cuenta:" se cambia a "Número de cta.", entonces la especificación de conversión anterior de la Figura 5 no puede identificar la etiqueta sensible y el estado se registra como el estado fallido. En tales escenarios, el módulo 212 de generación de consulta de contexto busca en una estructura del documento web, particularmente, en un documento web de HTML para identificar el conjunto de candidatos para la etiqueta sensible actualizada a través de una consulta de contexto. La consulta de contexto puede ser información de vecino capturada por el motor de creación de especificación de conversión durante el procedimiento de creación de especificación de conversión. El módulo 212 de generación de consulta de contexto genera la consulta de contexto para el documento web de HTML. Por ejemplo, un documento web de HTML 400 de la Figura 4 muestra la consulta de contexto dentro de una etiqueta de contexto. Ejemplos de consultas 600A-600C de contexto se muestran en las Figuras 6A-6C, respectivamente. En este ejemplo, la Figura 6A representa una consulta 600A de contexto literal, la Figura 6B representa una consulta 600B de contexto parametrizada y la Figura 6C representa la consulta 600C de contexto parametrizada con un ámbito horizontal aumentado. El segundo 'td' en la Figura 6C es

hermano del 'td' mostrado en la Figura 6B. Por lo tanto, la Figura 6C se denomina como la consulta de contexto parametrizada con el ámbito horizontal aumentado. El cambio en la consulta de contexto desde la consulta de contexto literal a parametrizada con ámbito aumentado se hace incrementalmente. La comparación de consulta de contexto literal es más preciso y genera un menor número de candidatos que la consulta de contexto parametrizada porque la probabilidad de comparar un mayor número de segmentos del documento web de HTML para la consulta de contexto parametrizada es mayor. El módulo 212 de generación de consulta de contexto puede permitir que el usuario final genere una consulta de contexto personalizada seleccionando o eliminando atributos de la consulta de contexto. Algunos de los atributos de la Figura 6C son ID, class, cellpadding y similares.

Además, el módulo 212 de generación de consulta de contexto ejecuta la consulta de contexto con el documento web de HTML y se determina el conjunto de candidatos que coinciden con la consulta de contexto. En un ejemplo, la ejecución de la consulta de contexto incluye la búsqueda de una lista de nodos que coincidan con un nodo superior de la consulta de contexto. Por ejemplo, en las Figuras 6A-6C, un nodo superior de consultas de contexto es "td". En este ejemplo, el módulo 212 de generación de consulta de contexto extrae una lista de nodos ejecutando la API de Modelo de Objeto de Documento (DOM) tal como getElementByTagName. Una vez que la lista de nodos se extrae a través de la API, a continuación el módulo 212 de generación de consulta de contexto compara la lista de nodos con los nodos de la consulta de contexto. El módulo 212 de generación de consulta de contexto a continuación considera los nodos que coinciden con la consulta de contexto como el conjunto de candidatos. El contenido marcado como "?" en la consulta de contexto se ignora en el procedimiento de comparación. Marcar algunas partes como "?" (ignorar) en la consulta de contexto es para ignorar cambios menores hechos al documento web de HTML sin impactar en el procedimiento de búsqueda. Las etapas de procedimiento anteriores se repiten para la determinación del conjunto de candidatos para los datos asociados con la etiqueta sensible.

Además, si la especificación de conversión falló debido a un cambio en la estructura del documento web (una etiqueta sensible y datos asociados con la etiqueta sensible no se cambian) que provoca que la interpretación de dirección falle, entonces el módulo 214 de detector de proximidad aumenta un ámbito de búsqueda para datos asociados con la etiqueta sensible manteniendo la etiqueta sensible como una referencia. Como se muestra en la Figura 7, la etiqueta sensible de 'Fecha de expiración' tiene únicamente un valor asociado con la misma. La Figura 7 muestra un documento 700 web de HTML en forma de árbol. Un documento 800 web de HTML en la Figura 8 se cambia o modifica para incorporar contenido nuevo y para mantener el diseño. La visualización de ambos de HTML en las Figuras 7 y 8 no se cambian debido a un diseño adicional pero sus representaciones (de árbol) internas se cambian. Si se crea una especificación de conversión para la etiqueta sensible de "Fecha de expiración" con dirección 'derecha', el motor de privacidad puede procesar y esconder los datos de 'Fecha de expiración' '05/06/2019' como se muestra en la Figura 7. Pero con la estructura cambiada del documento web de HTML y su representación de árbol (Figura 8), el motor de privacidad falla en la protección de los datos que corresponden a la etiqueta sensible 'Fecha de expiración' incluso aunque la etiqueta sensible y correspondientes datos no se cambian. En tal escenario, el módulo 214 de detector de proximidad aumenta el ámbito de búsqueda buscando en un fragmento de árbol más grande. De conformidad con la Figura 8, para obtener los datos para 'Fecha de expiración', se realiza una búsqueda de todo el árbol con raíz en el nodo 'Table' (destacado). Si la búsqueda se efectúa en el árbol con raíz en 'Table' entonces se encuentran dos candidatos que coinciden con 'Fecha de expiración' ('05/06/2019' y '05/06/2014') y uno de los dos candidatos pertenece a la etiqueta 'Fecha generada.'. Para seleccionar un candidato para 'Fecha de expiración' de entre las dos posibles fechas, el módulo 214 de detector de proximidad realiza un análisis de proximidad para 'Fecha de expiración'. Por ejemplo, el módulo 214 de detector de proximidad realiza un análisis de proximidad de múltiples formas. En un ejemplo, el módulo 214 de detector de proximidad realiza un análisis de proximidad comparando información de ubicación de ambos candidatos. En un ejemplo, información de ubicación (por ejemplo, XPath) para el nodo '05/06/2019' es /html/body/center/table/table[4]/table/tr/td/ e información de ubicación para el nodo '06/06/2014' es /html/body/center/table/table[4]/table[2]/tr/td. La información de ubicación de ambos candidatos difiere en un sexto nodo desde el principio y el índice (ubicación) con hermanos sugiere qué nodo de valor está más cercano a la etiqueta sensible 'Fecha de expiración'. El módulo 214 de detector de proximidad a continuación visualiza todos los candidatos coincidentes pero recomienda el más cercano. El usuario final puede seleccionar el más apropiado de entre una lista dada de candidatos coincidentes.

En otra realización, si la etiqueta sensible y datos asociados con la etiqueta sensible se cambian junto con su vecindad, el módulo 216 de transformación de respuesta extrae porciones del documento web procesado que está en una forma de un árbol en el que la especificación de conversión se aplica satisfactoriamente y el documento web de referencia. El módulo 216 de transformación de respuesta extrae porciones del árbol que incluyen una etiqueta sensible y correspondientes datos y todas las ramas que conectan la etiqueta sensible y correspondientes datos. Las porciones extraídas del documento web de referencia se denominan como árboles de entrada y las porciones del documento web se denominan como árboles objetivo. En un ejemplo, el módulo 216 de transformación de respuesta compara cada árbol de entrada con el documento web. El módulo 216 de transformación de respuesta a continuación extrae las porciones (árboles objetivo) del documento web que coinciden parcialmente con el árbol de entrada. El módulo 216 de transformación de respuesta a continuación obtiene un nodo de etiqueta sensible del árbol de entrada. Además, el módulo 216 de transformación de respuesta obtiene una consulta de contexto de la etiqueta sensible de la especificación de conversión. Para cada nodo hoja en el árbol objetivo, el módulo 216 de transformación de respuesta obtiene su información de vecindad de igual tamaño que la consulta de contexto. El módulo 216 de transformación de respuesta a continuación aplica operaciones de sustitución en la vecindad

extraída. Si la consulta de contexto de especificación de conversión y vecindad extraída es igual, entonces marcan el nodo hoja como una etiqueta sensible potencial. Si las operaciones de sustitución no generan la consulta de contexto requerida, el módulo 216 de transformación de respuesta aplica las operaciones de inserción en la vecindad extraída. El módulo 216 de transformación de respuesta a continuación compara la consulta de contexto de la especificación de conversión y vecindad extraída después de las operaciones de inserción. El módulo 216 de transformación de respuesta marca el nodo hoja como una etiqueta sensible potencial cuando existe una coincidencia entre la consulta de contexto de la especificación de conversión y vecindad extraída después de las operaciones de inserción. Si las operaciones de inserción no generan la consulta de contexto requerida, el módulo 216 de transformación de respuesta aplica las operaciones de borrado en la vecindad extraída. El módulo 216 de transformación de respuesta a continuación compara la consulta de contexto de la especificación de conversión y vecindad extraída después de las operaciones de borrado. El módulo 216 de transformación de respuesta marca el nodo hoja como una etiqueta sensible potencial cuando existe una coincidencia entre la consulta de contexto de la especificación de conversión y vecindad extraída después de las operaciones de borrado. En algunas realizaciones, se aplica una combinación de operaciones de sustitución, inserción y/o borrado en el árbol de entrada para obtener el árbol objetivo o se aplica una combinación de operaciones de sustitución, inserción y/o borrado en el árbol objetivo para obtener el árbol de entrada.

En un ejemplo, la Figura 9 ilustra un documento 900 web de HTML modificado de la Figura 4. 'Número de cuenta:' se ha cambiado a 'Número de cta.' y el nodo padre se cambia desde 'span' a 'div'. Una porción del documento 1000 web de HTML del documento 900 web de HTML de la Figura 9 se muestra en la Figura 10. Usando el documento web previsualizado como una referencia para comparar el documento web (una porción del que se muestra en la Figura 9), el módulo 216 de transformación de respuesta obtiene dos árboles, las Figuras 9 y 10, respectivamente, a partir del documento web. La Figura 10 tiene el mismo tamaño que la consulta de contexto mencionada en la Figura 6A o la Figura 6B. Si el módulo 216 de transformación de respuesta aplica operaciones de sustitución en la Figura 10 cambiando un nodo a 'span' desde 'div' y reemplazando el nodo hoja 'Número de cta.' con el parámetro desconocido '?', el código fuente de la Figura 10 se convierte en la consulta de contexto literal de la Figura 6A. A partir de la consulta de contexto literal, puede crearse la consulta de contexto parametrizada o la consulta de contexto con el ámbito horizontal aumentado. A continuación el módulo 216 de transformación de respuesta ejecuta la consulta de contexto nueva creada para determinar las etiquetas sensibles candidato potenciales.

En otro ejemplo, una porción modificada del código fuente de HTML de la Figura 4 se muestra en la Figura 11. La porción modificada del código fuente de HTML 1100 incluye el nodo padre 'span' de 'Número de cuenta' cambiado a 'Número de cta.'. El módulo 216 de transformación de respuesta realiza las operaciones de borrado en el código fuente de HTML original de la Figura 4 y produce el código fuente de HTML de la Figura 11. Para conseguir la consulta de contexto para la Figura 11, el módulo 216 de transformación de respuesta realiza una operación de inserción en la Figura 11 o una operación de borrado en la Figura 4. Además, el módulo 216 de transformación de respuesta ejecuta la consulta de contexto creada para determinar las etiquetas sensibles candidato potenciales. Adicionalmente, el módulo 216 de transformación de respuesta visualiza la lista de etiquetas sensibles potenciales para que el usuario final seleccione la etiqueta sensible de la lista de etiquetas sensibles. El módulo 216 de transformación de respuesta repite el procedimiento anterior para los datos asociados con las etiquetas sensibles.

Tras la determinación de las etiquetas sensibles actualizadas o datos asociados con las etiquetas sensibles, el módulo 218 de actualización de especificación de conversión actualiza la especificación de conversión con etiquetas sensibles actualizadas o datos asociados con las etiquetas sensibles. El módulo 218 de actualización de especificación de conversión habilita el motor de privacidad para el enmascaramiento de datos de los datos asociados con las etiquetas sensibles del documento web usando la especificación de conversión actualizada.

La Figura 12 es un diagrama 1200 de flujo que ilustra un procedimiento de habilitación de enmascaramiento de datos para documentos web, de acuerdo con una realización de la presente divulgación. En el bloque 1202, se reciben una especificación de conversión asociada con múltiples documentos web de referencia y documentos web de referencia previsualizados. Por ejemplo, la especificación de conversión es para datos asociados con etiquetas sensibles en los múltiples documentos web de referencia y los datos asociados con las etiquetas sensibles se determinan sensibles a base de al menos una regla de política definida por usuario. La especificación de conversión comprende información asociada con nodos de etiquetas sensibles en un documento web, información de vecino asociada con nodos de etiquetas sensibles, información de vecino de datos asociados con etiquetas sensibles e información asociada con una estructura del documento web. En el bloque 1204, se recibe un primer documento web procesado a base de la especificación de conversión, el primer documento web incluye una o más etiquetas sensibles y datos asociados con las etiquetas sensibles. En el bloque 1206, se determina un estado de la especificación de conversión tras el procesamiento del primer documento web. En algunas realizaciones, los múltiples documentos web de referencia previsualizados se analizan para determinar reglas de política definidas por usuario. La existencia de una etiqueta sensible en el primer documento web se determina a continuación usando las reglas de política definidas por usuario. Una alerta se envía a continuación al usuario cuando la etiqueta sensible existe en el primer documento web.

En el bloque 1208, se analizan el primer documento web y los múltiples documentos web de referencia previsualizados cuando el estado de la especificación de conversión es una etiqueta sensible no encontrada y/o datos asociados con una etiqueta sensible no encontrada. El primer documento web y los múltiples documentos web de referencia previsualizados se analizan para determinar un conjunto de candidatos para una etiqueta sensible y/o

un conjunto de candidatos para datos asociados con una etiqueta sensible en el primer documento web.

En un ejemplo, el estado de la especificación de conversión es la etiqueta sensible no encontrada cuando se actualiza la etiqueta sensible. En este ejemplo, el primer documento web se compara con los documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible en los documentos web de referencia previsualizados. Además, se determina una consulta de contexto para el nodo de etiqueta sensible usando información de vecino en la especificación de conversión para el nodo de etiqueta sensible. Adicionalmente, la consulta de contexto se ejecuta en el primer documento web para determinar el conjunto de candidatos para la etiqueta sensible actualizada en el primer documento web. En una realización, la consulta de contexto para el nodo de etiqueta sensible se compara con el primer documento web y el conjunto de candidatos para la etiqueta sensible actualizada se determinan a base de la comparación.

En otro ejemplo, el estado de la especificación de conversión son los datos asociados con la etiqueta sensible no encontrada cuando se actualizan los datos asociados con la etiqueta sensible. En este ejemplo, el documento web se compara con los documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible en los documentos web de referencia previsualizados. Además, una consulta de contexto para los datos asociados con el nodo de etiqueta sensible se determina usando información de vecino en la especificación de conversión para el nodo de etiqueta sensible. Adicionalmente, la consulta de contexto se ejecuta en el primer documento web para determinar el conjunto de candidatos para los datos asociados actualizados con la etiqueta sensible en el primer documento web.

En otro ejemplo más, el estado de la especificación de conversión son los datos asociados con la etiqueta sensible no encontrada cuando se actualiza la estructura del primer documento web con la etiqueta sensible y datos asociados con la etiqueta sensible no cambiada. En este ejemplo, el primer documento web se compara con los múltiples documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible. Además, se aumenta un conjunto de candidatos para los datos asociados con una etiqueta sensible que corresponde al nodo de etiqueta sensible aumentando un ámbito de búsqueda dentro del primer documento web. Adicionalmente, se realiza un análisis de proximidad en el conjunto de candidatos para los datos asociados con la etiqueta sensible para determinar el candidato correcto del conjunto de candidatos para los datos asociados con la etiqueta sensible.

En un ejemplo, el estado de la especificación de conversión son los datos asociados con la etiqueta sensible no encontrada cuando se actualiza una estructura del primer documento web con la etiqueta sensible y datos asociados con la etiqueta sensible cambiada. En este ejemplo, el primer documento web se compara con los documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible en los documentos web de referencia previsualizados. Además, se determina una consulta de contexto para el nodo de etiqueta sensible usando información en la especificación de conversión para el nodo de etiqueta sensible. Adicionalmente, se extrae información asociada con cada nodo en el primer documento web. Además, la consulta de contexto de la especificación de conversión se compara con la información extraída para determinar un conjunto de candidatos para la etiqueta sensible actualizada y/o los datos asociados actualizados con la etiqueta sensible.

En el bloque 1210, se actualiza la especificación de conversión tras la determinación del conjunto de candidatos para una etiqueta sensible y/o el conjunto de candidatos para datos asociados con la etiqueta sensible en el primer documento web. En el bloque 1212 se habilita el enmascaramiento de los datos asociados con las etiquetas sensibles en el primer documento web usando la especificación de conversión actualizada.

El orden en el que se describe el procedimiento o procedimientos no pretende que se interprete como una limitación, y cualquier número de los bloques de procedimiento descritos puede combinarse en cualquier orden para implementar el procedimiento 1200 o un procedimiento alternativo. Adicionalmente, pueden borrarse bloques individuales de los procedimientos sin alejarse del espíritu y ámbito del objeto descrito en el presente documento. Adicionalmente, el procedimiento 1200 puede implementarse en cualquier hardware, software, firmware o combinación adecuada de los mismos.

En una implementación, uno o más del procedimiento o procedimientos descritos en el presente documento pueden implementarse al menos en parte como instrucciones incorporadas en un medio de almacenamiento legible por ordenador no transitorio y ejecutable por uno o más dispositivos informáticos. En general, un procesador (por ejemplo un microprocesador) recibe instrucciones, desde un medio legible por ordenador no transitorio, por ejemplo, una memoria, y ejecuta esas instrucciones, realizando de este modo uno o más procedimiento o procedimientos, incluyendo uno o más del procedimiento o procedimientos descritos en el presente documento. Tales instrucciones pueden almacenarse y/o transmitirse usando cualquiera de una diversidad de medios legible por ordenador conocidos.

Sin embargo, debe apreciarse que el alcance de la protección se extiende a un programa de este tipo y además de un medio legible por ordenador que tiene un mensaje en el mismo; tales medios de almacenamiento legible por ordenador contienen medios de código de programa de implementación de una o más etapas del procedimiento, cuando el programa se ejecuta en un servidor o dispositivo móvil o cualquier dispositivo programable adecuado.

La descripción anterior se ha presentado con referencia a diversas realizaciones. Expertos en la materia y tecnología a la que pertenece esta aplicación aprecian que pueden practicarse alteraciones y cambios en las estructuras y procedimientos de operación descritos sin alejarse de manera significativa del alcance.

REIVINDICACIONES

1. Un procedimiento que comprende:

- 5 recibir una especificación de conversión asociada con múltiples documentos web de referencia y documentos web de referencia previsualizados, en el que la especificación de conversión es para datos asociados con etiquetas sensibles en los múltiples documentos web de referencia y en el que los datos asociados con las etiquetas sensibles se determinan sensibles a base de al menos una regla de política definida por usuario; recibir un primer documento web procesado a base de la especificación de conversión, en el que el primer documento web comprende al menos una etiqueta sensible y datos asociados con la al menos una etiqueta sensible;
- 10 determinar un estado de la especificación de conversión tras el procesamiento del primer documento web; analizar el primer documento web y los múltiples documentos web de referencia previsualizados cuando el estado de la especificación de conversión es al menos una de una etiqueta sensible no encontrada o datos asociados con una etiqueta sensible no encontrada, en el que el primer documento web y los múltiples documentos web de referencia previsualizados se analizan para determinar al menos uno de un conjunto de
- 15 candidatos para una etiqueta sensible y un conjunto de candidatos para datos asociados con una etiqueta sensible en el primer documento web; actualizar la especificación de conversión tras la determinación de al menos uno del conjunto de candidatos para una etiqueta sensible y el conjunto de candidatos para datos asociados con la etiqueta sensible en el primer documento web; y
- 20 habilitar el enmascaramiento de los datos asociados con la al menos una etiqueta sensible en el primer documento web usando la especificación de conversión actualizada.

2. El procedimiento de acuerdo con la reivindicación 1, en el que el estado de la especificación de conversión es la etiqueta sensible no encontrada cuando se actualiza la etiqueta sensible y en el que el estado de la especificación de conversión son los datos asociados con la etiqueta sensible no encontrada cuando al menos se actualiza uno de
- 25 los datos asociados con la etiqueta sensible, se actualiza una estructura del primer documento web con la etiqueta sensible y datos asociados con la etiqueta sensible no cambiada, y se actualiza una estructura del primer documento web con la etiqueta sensible y datos asociados con la etiqueta sensible cambiada.

3. El procedimiento de acuerdo con la reivindicación 2, en el que analizar el primer documento web y los múltiples documentos web de referencia previsualizados cuando el estado de la especificación de conversión es una etiqueta sensible no encontrada comprende:
- 30

- comparar el primer documento web con los múltiples documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible en los múltiples documentos web de referencia previsualizados; determinar una consulta de contexto para el nodo de etiqueta sensible usando información de vecino en la especificación de conversión para el nodo de etiqueta sensible; y
- 35 ejecutar la consulta de contexto en el primer documento web para determinar un conjunto de candidatos para la etiqueta sensible actualizada en el primer documento web.

4. El procedimiento de acuerdo con la reivindicación 3, en el que ejecutar la consulta de contexto en el primer documento web para determinar el conjunto de candidatos para la etiqueta sensible actualizada en el primer documento web, comprende:

- 40 comparar la consulta de contexto para el nodo de etiqueta sensible con el primer documento web; y determinar el conjunto de candidatos para la etiqueta sensible actualizada a base de la comparación.

5. El procedimiento de acuerdo con la reivindicación 2, en el que analizar el primer documento web y los múltiples documentos web de referencia previsualizados cuando el estado de la especificación de conversión son los datos asociados con la etiqueta sensible no encontrada comprende:

- 45 comparar el primer documento web con los múltiples documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible en los múltiples documentos web de referencia previsualizados; determinar una consulta de contexto para los datos asociados con el nodo de etiqueta sensible usando información de vecino en la especificación de conversión para el nodo de etiqueta sensible; y
- 50 ejecutar la consulta de contexto en el primer documento web para determinar un conjunto de candidatos para los datos asociados actualizados con una etiqueta sensible en el primer documento web.

6. El procedimiento de acuerdo con la reivindicación 2, en el que analizar el primer documento web y los múltiples documentos web de referencia previsualizados cuando se actualiza la estructura del primer documento web con la etiqueta sensible y datos asociados con la etiqueta sensible no cambiada comprende:

- 55 comparar el primer documento web con los múltiples documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible; determinar un conjunto de candidatos para los datos asociados con una etiqueta sensible que corresponde al nodo de etiqueta sensible aumentando un ámbito de búsqueda dentro del primer documento web; y realizar análisis de proximidad en el conjunto de candidatos para los datos asociados con la etiqueta sensible

para determinar el candidato correcto del conjunto de candidatos para los datos asociados con la etiqueta sensible.

7. El procedimiento de acuerdo con la reivindicación 2, en el que analizar el primer documento web y los múltiples documentos web de referencia previsualizados cuando se actualiza la estructura del primer documento web con la etiqueta sensible y datos asociados con la etiqueta sensible cambiada comprende:

comparar el primer documento web con los múltiples documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible en los múltiples documentos web de referencia previsualizados;
determinar una consulta de contexto para el nodo de etiqueta sensible usando información en la especificación de conversión para el nodo de etiqueta sensible;
extraer información asociada con cada nodo en el primer documento web; y
comparar la consulta de contexto de la especificación de conversión con la información extraída para determinar un conjunto de candidatos para al menos uno de la etiqueta sensible actualizada y los datos asociados actualizados con la etiqueta sensible.

8. El procedimiento de acuerdo con la reivindicación 1, que comprende adicionalmente:

analizar los múltiples documentos web de referencia previsualizados para determinar reglas de política definidas por usuario;
determinar la existencia de una etiqueta sensible en el primer documento web usando las reglas de política definidas por usuario; y
enviar una alerta cuando la etiqueta sensible existe en el primer documento web.

9. Un sistema (200) que comprende:

al menos un procesador (202); y
una memoria (204) acoplada comunicativamente al al menos un procesador (202), en el que la memoria (204) incluye un módulo (110) de habilitación de enmascaramiento de datos para:

recibir una especificación de conversión asociada con múltiples documentos web de referencia y documentos web de referencia previsualizados, en el que la especificación de conversión es para datos asociados con etiquetas sensibles en los múltiples documentos web de referencia y en el que los datos asociados con las etiquetas sensibles se determinan sensibles a base de al menos una regla de política definida por usuario;
recibir un primer documento web procesado a base de la especificación de conversión, en el que el primer documento web comprende al menos una etiqueta sensible y datos asociados con la al menos una etiqueta sensible;
determinar un estado de la especificación de conversión tras el procesamiento del primer documento web;
analizar el primer documento web y los múltiples documentos web de referencia previsualizados cuando el estado de la especificación de conversión es al menos una de una etiqueta sensible no encontrada o datos asociados con una etiqueta sensible no encontrada, en el que el primer documento web y los múltiples documentos web de referencia previsualizados se analizan para determinar al menos uno de un conjunto de candidatos para una etiqueta sensible y un conjunto de candidatos para datos asociados con una etiqueta sensible en el primer documento web;
actualizar la especificación de conversión tras la determinación de al menos uno del conjunto de candidatos para una etiqueta sensible y el conjunto de candidatos para datos asociados con la etiqueta sensible en el primer documento web; y
habilitar el enmascaramiento de los datos asociados con la al menos una etiqueta sensible en el primer documento web usando la especificación de conversión actualizada.

10. El sistema de acuerdo con la reivindicación 9, en el que el estado de la especificación de conversión es la etiqueta sensible no encontrada cuando se actualiza la etiqueta sensible y en el que el estado de la especificación de conversión son los datos asociados con la etiqueta sensible no encontrada cuando al menos se actualiza uno de los datos asociados con la etiqueta sensible, se actualiza una estructura del primer documento web con la etiqueta sensible y datos asociados con la etiqueta sensible no cambiada y se actualiza una estructura del primer documento web con la etiqueta sensible y datos asociados con la etiqueta sensible cambiada.

11. El sistema de acuerdo con la reivindicación 10, en el que el módulo (110) de habilitación de enmascaramiento de datos se configura para:

comparar el primer documento web con los múltiples documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible en los múltiples documentos web de referencia previsualizados;
determinar una consulta de contexto para el nodo de etiqueta sensible usando información de vecino en la especificación de conversión para el nodo de etiqueta sensible; y
ejecutar la consulta de contexto en el primer documento web para determinar un conjunto de candidatos para la etiqueta sensible actualizada en el primer documento web.

12. El sistema de acuerdo con la reivindicación 11, en el que el módulo (110) de habilitación de enmascaramiento de datos se configura para:

comparar la consulta de contexto para el nodo de etiqueta sensible con el primer documento web; y
determinar el conjunto de candidatos para la etiqueta sensible actualizada a base de la comparación.

13. El sistema de acuerdo con la reivindicación 10, en el que el módulo (110) de habilitación de enmascaramiento de datos se configura para:

- 5 comparar el primer documento web con los múltiples documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible en los múltiples documentos web de referencia previsualizados;
- determinar una consulta de contexto para los datos asociados con el nodo de etiqueta sensible usando información de vecino en la especificación de conversión para el nodo de etiqueta sensible; y
- 10 ejecutar la consulta de contexto en el primer documento web para determinar un conjunto de candidatos para los datos asociados actualizados con una etiqueta sensible en el primer documento web.

14. El sistema de acuerdo con la reivindicación 10, en el que el módulo (110) de habilitación de enmascaramiento de datos se configura para:

- comparar el primer documento web con los múltiples documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible;
- 15 determinar un conjunto de candidatos para los datos asociados con una etiqueta sensible que corresponde al nodo de etiqueta sensible aumentando un ámbito de búsqueda dentro del primer documento web; y
- realizar análisis de proximidad en el conjunto de candidatos para los datos asociados con la etiqueta sensible para determinar el candidato correcto del conjunto de candidatos para los datos asociados con la etiqueta sensible.

20 15. El sistema de acuerdo con la reivindicación 10, en el que el módulo (110) de habilitación de enmascaramiento de datos se configura para:

- comparar el primer documento web con los múltiples documentos web de referencia previsualizados para obtener un nodo de etiqueta sensible en los múltiples documentos web de referencia previsualizados;
- 25 determinar una consulta de contexto para el nodo de etiqueta sensible usando información en la especificación de conversión para el nodo de etiqueta sensible;
- extraer información asociada con cada nodo en el primer documento web; y
- comparar la consulta de contexto de la especificación de conversión con la información extraída para determinar un conjunto de candidatos para al menos uno de la etiqueta sensible actualizada y los datos asociados actualizados con la etiqueta sensible.
- 30

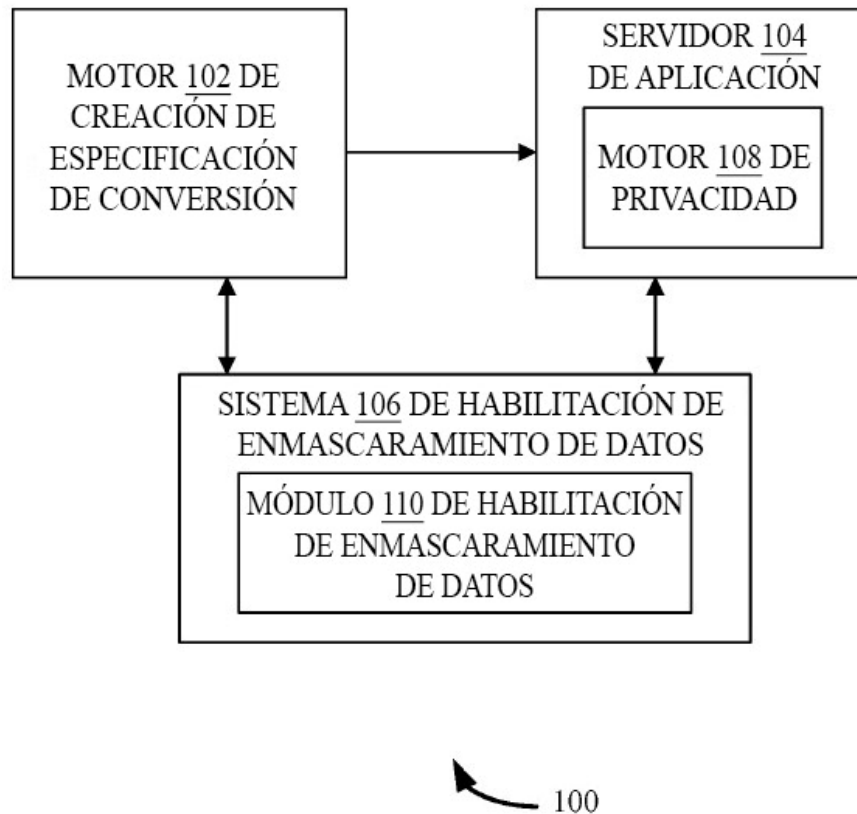


FIG. 1

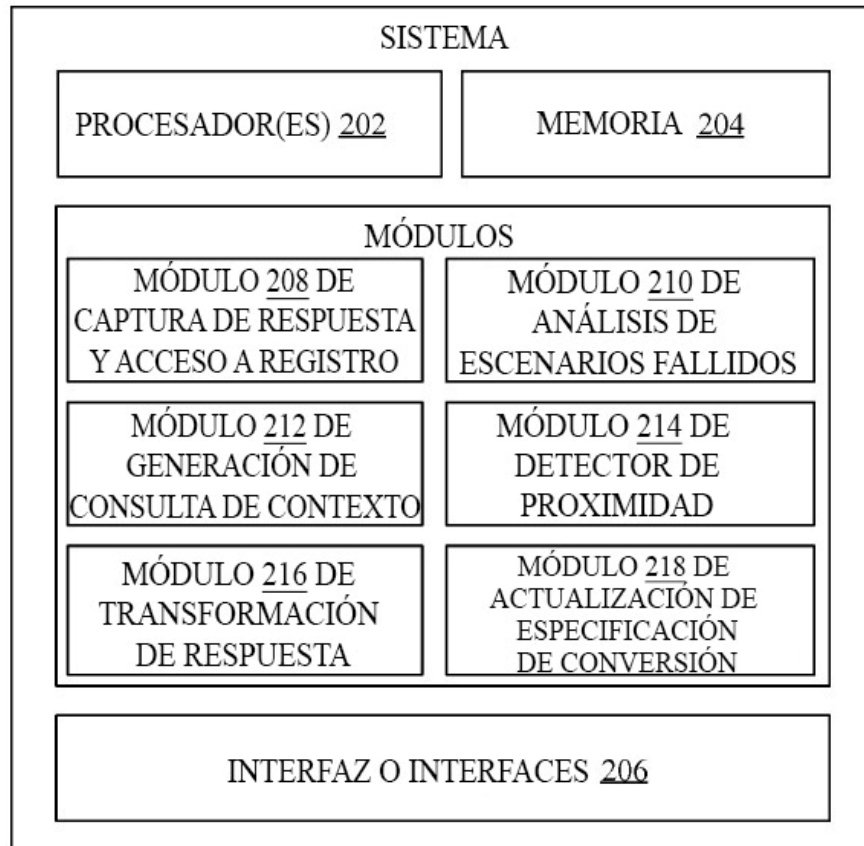


FIG. 2  200

Detalles de resumen de cuenta	
Tipo de tarjeta:	MasterCard
Número de cuenta:	<u>5555 6666 2222 1111</u>
Estado:	NORMAL/ACTIVA
Total debido:	\$ 47,74
Mínimo debido:	\$ 47,74
Cantidad vencida:	\$ 14,32
Cantidad de límite de crédito:	\$ 5.000,00
Fecha de vencimiento:	21/10/2010

FIG. 3  300

```

<tr class="subgroupBackGround">
  <td nowrap class="labelText"><span class="xq">
    Número de cuenta:</span></td>
  <td align="right" nowrap><table
    id="_id468_xc" class="xlu p_AFReadOnly"
    cellpadding="0" cellspacing="0" border="0"
    summary="">
    <tr>
      <td class="xu" nowrap></td>
      <td valign="top" nowrap class="AFContentCell"><div
        id="_id468" class="labelTextBold"
        style="font-weight: bold;">5555 6666
        2222 1111</div></td>
    </tr>
    <tr>
      <td></td>
      <td class="AFComponentMessageCell"><span
        class="x82"></span></td>
    </tr>
  </table></td>
</tr>

```

FIG. 4

400


```

<config
  hash=
    "9653d876a0c2ec7154ac96e365e67c6afc873c41f192f126b3b1a57412c6c29257225
    e3fe5f83f9760ac1a7e38d3ca6a74c52acf09899dddbaf8cf4cd862e507">
  <rules storyboard="page_identifier_0012020">
    <rule direction="down" labellength="14" type="text">
      <lpattern>Número\s*Cuenta\s|Número\s*Cuenta\s*:\s*
    </lpattern>
    <vpattern>
      [0-9](4)\s*[0-9](4)\s*[0-9](4)\s*[0-9](4)\s*[0-9](4)\s*[0-9](13)
    </vpattern>
    <attrs />
    <context>
      <td nowrap class="labelText"><span class="xq"?
    </span></td>
    </context>
    <action method="substitution" type="mask">
      <params>all,all,X</params>
    </action>
  </rule>
</rules>

```

FIG. 5

500

```
<td nowrap class="labelText"><span class="xq"?></span></td>
```

FIG. 6A



600A

```
<td nowrap class="?"><span class="?"></span></td>
```

FIG. 6B



600B

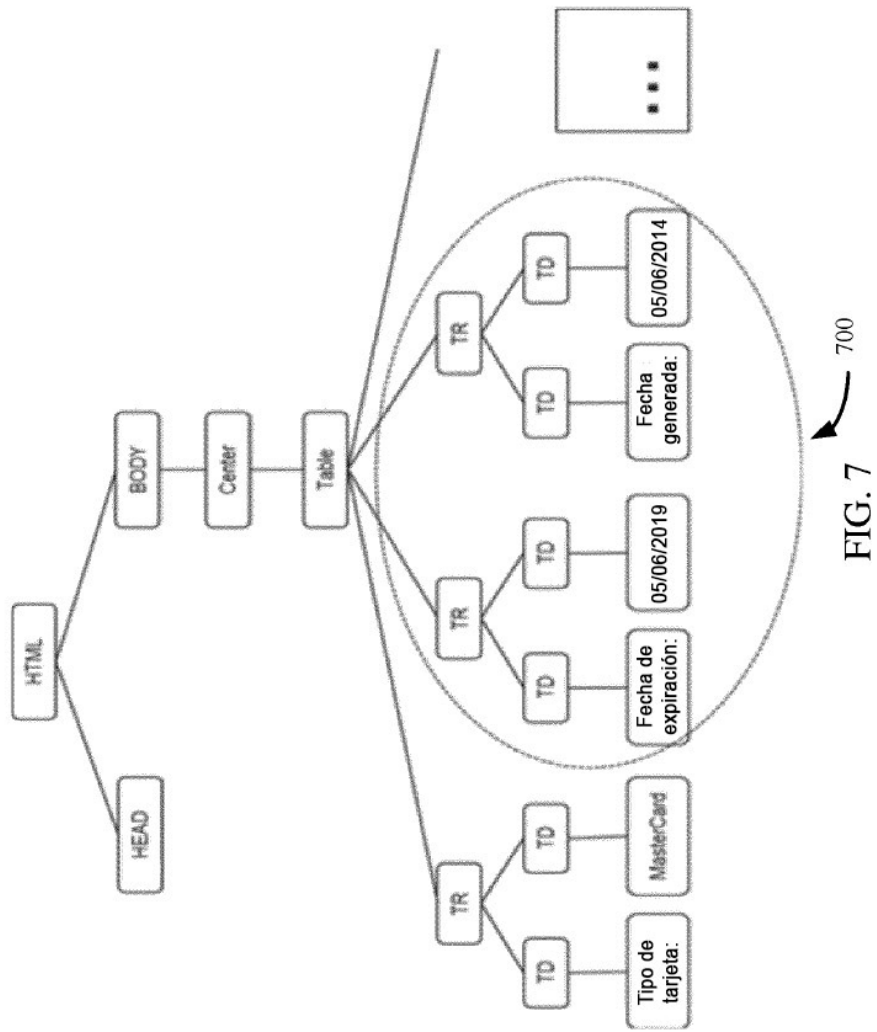
```

<td nowrap class="labelText"><span class="xq">
??</span></td>
<td align="right" nowrap><table
id="_id468_xc" class="xlu p_AFReadOnly"
cellpadding="0" cellspacing="0" border="0"
summary="">
<tr>
<td class="xu" nowrap></td>
<td valign="top" nowrap class="AFContentCell"><div
id="_id468" class="labelTextBold"
style="font-weight: bold;"??
|</div></td>
</tr>
<tr>
<td></td>
<td class="AFComponentMessageCell"><span
class="x8z"></span></td>
</tr>
</table></td>
</td></td>

```

FIG. 6C

600C



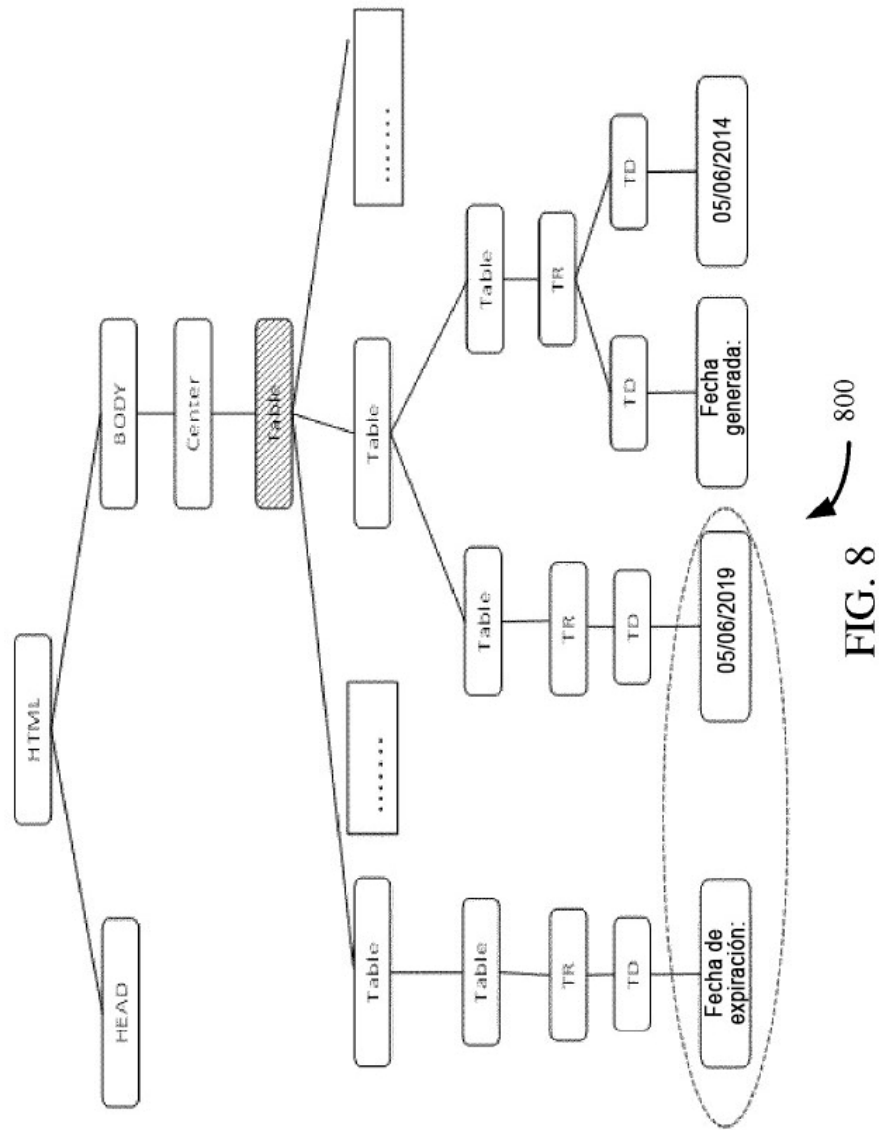


FIG. 8

```

<td nowrap class="labelText"><div class="xq">Número de cta.</div></td>
<td align="right" nowrap>
<table id="_id468_xc" class="xln p_AFReadOnly" cellpadding="0"
cellspacing="0" border="0" summary="0">
  <tr>
    <td class="xu" nowrap></td>
    <td nowrap valign="top" class="AFContentCell"><div id=
      "_id468" class="labelTextBold" style="font-weight:bold;"
    ></div></td>
  </tr>
  <tr>
    <td></td>
    <td class="AFComponentMessageCell"><div class="x82">5555
      6666 2222 1111</div></td>
  </tr>
</table>

```

FIG. 9

900

```
<td nowrap class="labelText"><div class="xq">Número de cta.</div></td>
```

FIG. 10

```
<td nowrap class="labelText">Número de cta.</td>
```

FIG. 11

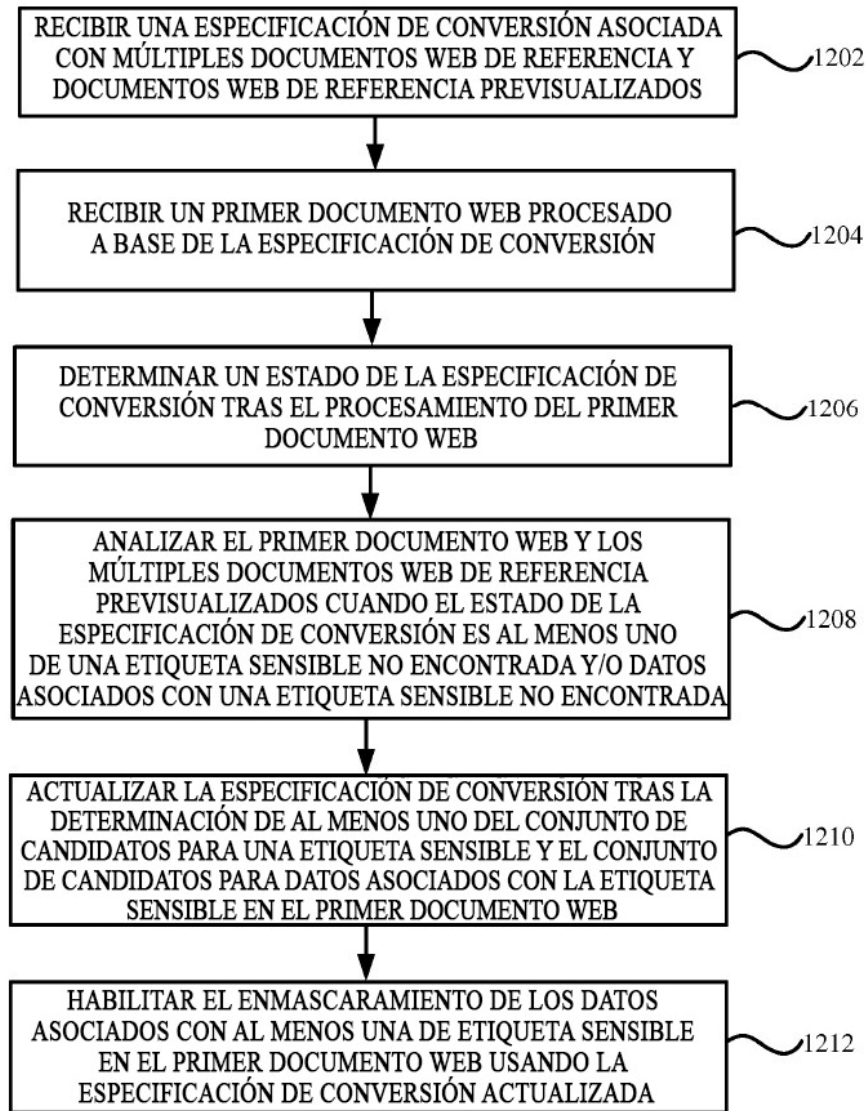


FIG. 12 1200