(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 763 825**

(51) Int. CI.:

***G06F 16/22*** (2009.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **09.04.2018 E 18166222 (2)**

(97) Fecha y número de publicación de la concesión europea: **16.10.2019 EP 3388969**

(54) Título: **Sistema de búsqueda**

(30) Prioridad:

**13.04.2017 EP 17166553**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
**01.06.2020**

(73) Titular/es:

**DSWISS AG (100.0%)**
**Badenerstrasse 329**
**8003 Zürich, CH**

(72) Inventor/es:

**GERMANN, FABIO;**
**TSCHANNEN, MICHAEL y**
**PAGANONI, SERGIO**

(74) Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

## DESCRIPCIÓN

Sistema de búsqueda

**Campo técnico**

La presente invención se refiere a un sistema de búsqueda que comprende un dispositivo de almacenamiento, un servidor de búsqueda, en el que el dispositivo de almacenamiento comprende un almacenamiento de archivos, un almacenamiento de índices, una base de datos de claves de documentos, una base de datos de índices y una base de datos de usuarios/claves.

**Técnica anterior**

Los servicios de almacenamiento pueden proporcionarse de manera local o como un servicio en la nube. Los servicios de almacenamiento almacenan, en general, datos de documentos cargados por los usuarios y permiten a los usuarios y a los grupos de usuarios, a quienes los usuarios les dan permiso, compartir los datos del documento. Los sistemas de gestión de documentos (DMS) a menudo se integran dentro de dichos servicios de almacenamiento y deberían proporcionar una búsqueda de texto completo que busca datos de documentos incluyendo palabras clave.

Las consideraciones de seguridad solicitan que los datos se almacenen como datos cifrados. A continuación, será necesario implementar estrategias de búsqueda en estos datos cifrados. En el marco de trabajo de los servicios en la nube, una pluralidad de usuarios comparten servidores y espacio de almacenamiento. Por lo tanto, el almacenamiento que incluye los datos del documento de un usuario puede ser parte de una consulta de búsqueda, incluso si los datos del usuario no son el objetivo de dicha búsqueda. A continuación es necesario tomar medidas para evitar que se filtre información confidencial. Una de tales propuestas con las características del preámbulo de la reivindicación 1 se describe en el documento US 2016/0 299 924 A1. Dicho documento de la técnica anterior proporciona, en el entorno típico de uso de internet, funciones de búsqueda de datos en el almacenamiento de proveedores de servicios en la nube con los datos cifrados.

Un sistema de búsqueda de acuerdo con el documento US 2016/0 299 924 A1 comprende: un dispositivo de almacenamiento configurado para almacenar un documento y un índice que se cifran con un formato de búsqueda; y un servidor de búsqueda configurado para buscar datos almacenados en el dispositivo de almacenamiento, en el que el dispositivo de almacenamiento está configurado para almacenar: una base de datos de índices que almacena un índice cifrado, que es un índice cifrado con una clave de índice, y una clave de índice cifrada, que es la clave de índice cifrada con una clave de usuario, asociada con el índice cifrado; y una base de datos de documentos que almacena un documento cifrado, que es el documento cifrado con una clave de documento, y una clave de documento cifrada, que es la clave de documento cifrada con la clave de usuario, asociada con el documento cifrado, en el que el servidor de búsqueda está configurado para extraer un término para buscar a partir de una consulta de búsqueda solicitada, descifrar la clave de índice cifrada con una clave de usuario que pertenece a un usuario que solicita la consulta de búsqueda, cifrar el término extraído con la clave de índice descifrada sin almacenar la clave de índice descifrada en un dispositivo de almacenamiento no transitorio, buscar la base de datos de índices con el índice cifrado usando el término cifrado y enviar un resultado de búsqueda a un terminal.

El documento US 2016 / 0 299 924 A1 desvela además un sistema para añadir documentos a una base de datos de documentos cifrados existente, en el que el servidor de búsqueda está configurado para extraer un término del documento, que se solicitó que se registre, cifrar el término extraído del documento con la clave de índice, cifrar la clave de índice con una clave de usuario que pertenezca a un usuario que solicita registrar el documento a registrar, registrar el término cifrado extraído del documento y la clave de índice cifrada en la base de datos de índices, cifrar el documento a registrar con la clave de documento, cifrar la clave de documento con la clave de usuario que pertenece al usuario que solicita registrar el documento, y registrar el documento cifrado y la clave de documento cifrado en la base de datos de documentos.

El documento US 2016/330 180 se refiere a una base de datos cifrada de conocimiento cero, donde las claves de cifrado se retienen del sistema que almacena la base de datos.

El documento US 2016/379 009 desvela un índice de búsqueda personal mejorado de privacidad donde la base de datos que almacena los documentos, especialmente en la nube, almacena documentos cifrados y es capaz de usar consultas de búsqueda opacas para acceder a los documentos cifrados.

El documento US 2017/091 475 se refiere a un método para la búsqueda de datos cifrados en un servidor no confiable a través de un índice de archivo local con valores de búsqueda que se producen en al menos un archivo cifrado en dicho servidor no confiable.

El documento US 2008/059414 desvela una búsqueda de datos cifrados con un cálculo de un valor de indexación que se usa dentro de una estructura de indexación para buscar un artículo que corresponde al valor de indexación

basado en el E-valor correspondiente al artículo descubierto.

El documento US 2005/004924 desvela un sistema de base de datos que comprende un índice de cifrado y un componente seguro capaz de manipular dicho índice de cifrado.

**Sumario de la invención**

Los documentos de la técnica anterior proporcionan un enfoque seguro para buscar documentos cifrados. Sin embargo, este enfoque solo permite extraer/buscar palabras clave de un documento, pero evita las "funciones de búsqueda avanzada", incluidas, pero no limitadas a, las búsquedas semánticas. Uno de los principales objetivos de la presente invención es proporcionar un sistema de búsqueda que comprenda una base de datos cifrada dentro de la cual es posible realizar enfoques/características de búsqueda arbitrarias. La búsqueda semántica considera, entre otros, diversos puntos que incluyen el contexto de búsqueda, localización, intención, variación de palabras, sinónimos, consultas generalizadas y especializadas, coincidencia de conceptos y consultas en lenguaje natural para proporcionar resultados de búsqueda relevantes. Pueden encontrarse más elementos relacionados con la búsqueda semántica en una entrada de artículo de Tony John "What is Semantic Search and how it works with Google search" del 15 de marzo de 2012 en Techulator en http://www.techulator.com/resources/ 5933-What-Semantic- Search.aspx. La principal diferencia entre los documentos de la técnica anterior y la presente invención es la forma en que se realiza una búsqueda: mientras que los documentos de la técnica anterior buscan un término cifrado en un índice cifrado, la presente invención busca un término de texto plano en un índice parcialmente descifrado. La presente invención describe un enfoque de encriptación de índice que técnicamente permite una característica de búsqueda avanzada como es una búsqueda semántica. La invención se define por las características de las reivindicaciones independientes. Otras realizaciones son el tema de las reivindicaciones dependientes.

A este respecto un objetivo adicional de la presente invención es proporcionar un sistema de actualización de la base de datos cifrada y otros elementos de base de datos mientras se mantiene el enfoque de búsqueda semántica.

Un sistema de búsqueda de acuerdo con la invención comprende un dispositivo de almacenamiento y un servidor de búsqueda configurado para buscar datos almacenados en el dispositivo de almacenamiento. El dispositivo de almacenamiento comprende un almacenamiento de archivos configurado para almacenar al menos un documento, en el que el dispositivo de almacenamiento comprende: un almacenamiento de archivos configurado para almacenar al menos un documento, un almacenamiento de índices configurado para almacenar al menos un índice, cada uno de estos asociados o relacionados con al menos un documento, una base de datos de claves de documento configurada para almacenar al menos una clave de documento asociada a un documento en el almacenamiento de archivos y en la que la clave de documento se usa para cifrar y descifrar el documento asociado, una base de datos de índices configurada para almacenar al menos una clave de índice, asociada a el al menos un índice en el almacenamiento de índices y en la que la clave de índice se usa para cifrar y descifrar el índice asociado y una base de datos de clave de usuario configurada para almacenar al menos una clave de usuario. Cada clave de usuario está asociada a al menos una clave de índice y al menos a una clave de documento a través de usar la clave de usuario respectiva para cifrar y descifrar las claves de índice asociadas en la base de datos de índices, así como las claves de documento asociadas en la base de datos de documentos.

A continuación, el servidor de búsqueda está configurado para extraer uno o más términos de búsqueda de una consulta de búsqueda solicitada, para descifrar la clave de índice cifrada con la clave de usuario que pertenece al usuario que solicita la consulta de búsqueda, para buscar una o múltiples secuencias/árboles de nodos en la base de datos de índices, descifrar jerárquicamente los nodos y enviar un resultado de búsqueda a un terminal.

Por lo tanto, el índice cifrado comprende al menos un árbol de índices que comprende una secuencia de nodos interconectados estando cada uno cifrado de manera individual, y que comprende la información de referencia de índice, así como información relevante para nodos secundarios interconectados.

El almacenamiento de índices, por lo general, almacena una pluralidad de índices, estando cada uno de estos asociado a al menos un documento, es decir, cada índice es capaz de almacenar información acerca de uno a muchos documentos.

La presente solicitud se usa la expresión "cifrado de nodo de árbol". El cifrado de nodo de árbol se refiere a un árbol que se cifra sobre una base de nodo, es decir, cada nodo de árbol se cifra individualmente, con un algoritmo de cifrado arbitrario. Una definición de característica adicional está relacionada con el "cifrado basado en nodos". Esta característica está relacionada con "la información cifrada de tal manera que es posible el descifrado basado en nodos". La información relevante sobre los nodos secundarios (y los nodos potencialmente vecinos) se almacena por lo tanto en la información del nodo primario. Con esto, es posible recorrer un árbol que consiste en nodos cifrados, sin la necesidad de descifrar el árbol completo. Dado que la información sobre la siguiente etapa de recorrido está contenida en el nodo primario, el sistema puede descifrar solo los nodos relevantes, con el fin de realizar un recorrido del árbol. Dicho esto, un proceso de búsqueda solo revelará tanta información como sea necesaria para completar la consulta, y no expondrá toda la información almacenada en el índice.

Como alternativa, la información puede mantenerse también en los nodos secundarios, lo que conduciría al "descifrado basado en niveles" en lugar del descifrado basado en nodos o una combinación de los mismos. Del mismo modo que el descifrado basado en nodos se refiere al descifrado de nodos secundarios, el descifrado basado en niveles se relaciona con el descifrado lateral. Técnicamente dicho, la cantidad de información descifrada es la misma en ambos casos, lo que cambia es el orden en que se recorre y se descifra el árbol.

Una hoja es un nodo secundario en el árbol que proporciona la información solicitada. Una hoja de este tipo también puede apuntar a otro índice diferente en lugar de a un documento. También puede hacer referencia a parte de un documento coincidente. Por supuesto, cada índice puede tener varios árboles.

Un nodo está representado por una estructura de datos arbitraria, que puede, por ejemplo, (pero no limitado a) ser una matriz, una lista, un hashmap, o cualquier otra estructura de datos conocida en tecnología de la información. El árbol se define por la presente como una forma de interconectar dichas partes de los datos.

Entonces, es una ventaja cuando cada nodo comprende una o más referencias en nodos y/o hojas conectados, ya que la etapa de descifrado durante una búsqueda solo está relacionada con un número de estos nodos que se descifran sobre una base individual, en comparación con un término de búsqueda específico, y si la comparación es exitosa, la información sobre los vecinos interconectados o nodos secundarios contenidos dentro del nodo descifrado se usa para encontrar el siguiente nodo índice a descifrar. Si se completa la búsqueda, los documentos mencionados en el último nodo, que son una hoja para una coincidencia positiva, se reenvían como resultado de la respuesta para la consulta de búsqueda.

Para mejorar la calidad de una búsqueda, pueden consultarse otros índices y subíndices para obtener información adicional para precisar y enriquecer la búsqueda. Esto también se aplica al proceso de índice, que podría actualizar más de un índice para un documento.

El sistema de búsqueda comprende una estructura de índices que permite un descifrado basado en nodos y niveles, permitiendo de este modo ventajas de seguridad y privacidad, que solo tiene que descifrar partes relevantes (en lugar de descifrar todo el índice) del índice con el fin de ejecutar una búsqueda, en la que en una estructura de índices de este tipo cada nodo de árbol contiene información de índice y referencias a otros nodos, nodos secundarios y nodos de hoja que contienen referencias a un o unos documentos coincidentes, otros índices y metadatos de búsqueda.

El resultado de la búsqueda, como se define en el presente contexto, comprende, en la realización más sencilla, documentos como tales. Ya que los detalles interiores de la presentación de un resultado de búsqueda pueden diferir de un producto a otro, y especialmente pueden incluir punteros a documentos en el sistema de archivos (es decir, lo que el usuario en general quiere), además de metadatos de soporte que pueden mejorar la visualización de resultados de búsqueda, tal como una posición de resultado de búsqueda en el documento. El resultado de la búsqueda puede proporcionar un documento coincidente en el almacenamiento de archivos, así como punteros a otros índices y/o a metadatos de búsqueda tales como las partes del índice.

El sistema comprende preferentemente una estructura de índices que permite un descifrado basado en nodos y niveles. Una estructura de índices de este tipo se conoce en la técnica. La presente invención funciona con varias estructuras de índice, que son la base para que la invención funcione correctamente, es decir, almacenar información en un árbol, con información sobre los nodos secundarios que se almacenan en el nodo principal. Por ejemplo, Lucene está trabajando de este modo: véase https://lucene.apache.org/. A continuación, la invención permite descifrar solo partes relevantes del índice con el fin de ejecutar una búsqueda, en la que cada nodo de árbol contiene la información de índice y las referencias a otros nodos, nodos secundarios y nodos de hoja que contienen referencias a documentos coincidentes, otros índices y metadatos de búsqueda.

Otras realizaciones de la invención se establecen en las reivindicaciones dependientes.

**Breve descripción de los dibujos**

Las realizaciones preferidas de la invención se describen a continuación haciendo referencias a los dibujos, que son con el fin de ilustrar las presentes realizaciones preferidas de la invención y no con el fin de limitar la misma. En los dibujos,

la figura 1: muestra un diagrama de bloques de un sistema de almacenamiento y de búsqueda de documentos de acuerdo con una realización de la invención;
la figura 2: muestra el sistema de almacenamiento y de búsqueda de acuerdo con la figura 1;
la figura 3: muestra el enfoque de procesamiento de documentos almacenados y accedidos por dos usuarios A y B;
la figura 4: muestra el enfoque de procesamiento de documentos almacenados y accedidos por dos grupos de usuarios A, B así como A, C;

la figura 5: muestra un árbol de decisión del sistema de almacenamiento y de búsqueda de documentos de acuerdo con la figura 1;
la figura 6: muestra un árbol de nodos del índice accedido por el motor de búsqueda y cómo el sistema de almacenamiento y de búsqueda de documentos de acuerdo con la figura 1 aplica el descifrado basado en nodos de acuerdo con la figura 5;
la figura 7: muestra una aplicación del iniciador de la figura 1;
la figura 8A y B: muestra un diagrama de flujo del proceso de indexación para un nuevo documento en el sistema de almacenamiento y de búsqueda de acuerdo con la figura 1 en dos hojas conectadas; y
la figura 9: muestra un diagrama de flujo del proceso de búsqueda en el sistema de almacenamiento y de búsqueda de acuerdo con la figura 1.

**Descripción de las realizaciones preferidas**

La figura 1 muestra un diagrama de bloques de un sistema de almacenamiento y de búsqueda de acuerdo con una realización de la invención. El sistema de almacenamiento y de búsqueda de documentos 200 comprende una base de datos de usuario y claves 205, una base de datos de documentos 206, una base de datos de índices 207, un almacenamiento de archivos 208 y un almacenamiento de índices 209. El sistema de almacenamiento y de búsqueda de documentos 200 se controla a través de un servidor de búsqueda 203 y un servidor distribuidor 204. El sistema de almacenamiento y de búsqueda 200 puede impulsarse por un iniciador 201 que puede ser un terminal de usuario 210 o un servidor de aplicaciones 202.

La figura 2 muestra el sistema de almacenamiento y de búsqueda 300 de acuerdo con la figura 1, referenciado como 200 en la misma. Sin embargo, solo se atribuyen números de referencia parcialmente similares, ya que son posibles otros enfoques del terminal de usuario 210. El sistema de almacenamiento y de búsqueda de documentos 300 comprende un sistema informático conocido que tiene una base de datos de usuarios y claves 302 que almacena una pluralidad de claves de usuario cifradas 303 (obsérvese que las claves de usuario también pueden almacenarse sin cifrar, lo que no tiene ningún efecto en la presente invención). Se proporciona una base de datos de documentos 304 que tiene una pluralidad de claves de documento cifradas 305 que apuntan a documentos cifrados 309 en el almacenamiento de archivos 308. Finalmente, una base de datos de índices 306 comprende una pluralidad de claves de índice cifradas 307 que apuntan a índices cifrados 311 en el almacenamiento de índices 310. En la parte de almacenamiento adicional del sistema informático, el almacenamiento de archivos 308 comprende y almacena una pluralidad de documentos 309, que pueden tener varios formatos. Los documentos comprenden al menos parcialmente una parte de búsqueda de palabras o palabras clave. El almacenamiento de índices 310 comprende al menos un índice 311.

Un usuario puede ponerse en contacto con el sistema de almacenamiento y de búsqueda 300 con un terminal de usuario 301. Esto puede ser un teclado local y una pantalla o un inicio de sesión remoto u otro medio que permita registrar una solicitud de búsqueda.

La figura 3 muestra el enfoque para manejar los documentos que se almacenan en un sistema de almacenamiento y de búsqueda 200 o 300 de acuerdo con la figura 1 o 2 y accedido por dos usuarios A y B, que usan las instancias 402/401 o un terminal 210 (figura 1/7), respectivamente.

En la siguiente explicación, el documento se procesa o bien para permitir que una sola persona o usuario, como se explica en conexión con la figura 3, o bien unas personas como miembros de un grupo de usuarios, como se explica en relación con la figura 4, busquen y recuperen un documento.

En la vista de las claves/base de datos de la figura 3, hay unos usuarios 400 y un material de clave asociado 410. En aras de la simplicidad, el dibujo se relaciona con dos usuarios A y B que acceden al sistema con los terminales 401 y 402, respectivamente. Antes de interactuar con el sistema de búsqueda 200 o 300, el cliente (en este caso el usuario A y el usuario B) se autentica así mismo con su contraseña o con un mecanismo de autenticación alternativo para desbloquear su material clave. A continuación, el usuario A tiene acceso a la clave de usuario_a y el usuario B tiene acceso a la clave de usuario_b. Esto puede realizarse usando el terminal con un navegador web u otro sistema informático que actúe en nombre del usuario. En otras palabras, el usuario A y el usuario B acceden a la base de datos de usuario/claves, se autentican a sí mismos como usuario_a o usuario_b con su contraseña habitual más eventualmente una autenticación de 2 o 3 factores y desbloquean su clave de usuario_a y clave de usuario_b, respectivamente. Una vez completado esta etapa con éxito, el usuario A y el usuario B reciben un testigo respectivo y/o un secreto. El testigo y/o un secreto no se materializan como tales en la figura 3 sino que se asocia a las flechas de desbloqueo 421 y 422, respectivamente.

La base de datos de índices 306 comprende una pluralidad de claves de índice cifradas 307 y los índices cifrados autorizados asociados 430 que se almacenan en el almacenamiento de índices. El ejemplo con dos usuarios usuario_a y usuario_b proporciona la siguiente situación. El testigo 421 del usuario_a es capaz de acceder y desbloquear la clave de usuario_a, que es capaz de desbloquear la clave de índice_a, perteneciente al índice 430 índice_a para el usuario_a así como la clave de índice_c, perteneciente al índice_c almacenado para más de un usuario (grupo de usuarios), en este caso para el usuario_a y el usuario_b. Por supuesto, la base de datos de

índices 306 comprende más claves de índice 307 para los índices correspondientes en el almacenamiento de índices 430, para el usuario_a, el usuario_b y/u otros usuarios, así como para grupos de usuarios.

El almacenamiento de índices 310 comprende una pluralidad de índices 430. El ejemplo con dos usuarios A y B proporciona la siguiente situación. La clave de usuario_a 421 del usuario A es capaz de acceder y desbloquear las claves de índice (clave de índice_a y clave de índice_c) proporcionando acceso al índice a, que pertenece a los documentos almacenados solo para el usuario A, así como al índice c, que pertenece a los documentos almacenados (en el presente ejemplo) para un grupo compuesto por el usuario A y el usuario B.

La figura 4 muestra el enfoque para manejar los grupos de usuarios usuario_a y usuario_b.; la base de datos de índices se complementa con una base de datos de grupo, donde los grupos pueden contener uno o más usuarios. En el presente ejemplo, el usuario A es miembro tanto de un "grupo implícito" (Grupo 1, que es solo el "usuario_a", que tiene acceso a su índice personal) como de un grupo explícito el Grupo 3, que permite el acceso a un índice compartido junto con el usuario_b. Habrá, por supuesto, grupos para los que el usuario_a no tiene autorización.

Cada grupo explícito e implícito tiene básicamente un índice, y una clave de índice que pertenece a este índice. En otras palabras, un "grupo" (que consiste en al menos un usuario) tiene acceso a al menos un documento, que está cubierto en un índice distinto al que puede acceder el "grupo", a través de la clave de índice. Por lo tanto, dicha clave de índice puede desbloquearse por todas las claves de usuario de los usuarios que son miembros de un grupo.

La figura 5 muestra un árbol 500 del sistema de almacenamiento y de búsqueda 200 o 300, y describe cómo se almacena la información de índice. El árbol comprende una pluralidad de nodos de árbol 502. Cada nodo de árbol comprende referencias a nodos vecinos y/o a nodos secundarios, visualizados por las flechas de conexión de nodo 503 y 513. Esto se refiere a las referencias a todos los nodos secundarios y a la información relevante para elegir el siguiente nodo. Una cadena de búsqueda comprende varias palabras clave que identificarán una palabra clave después de la otra y, por lo tanto, seguirán la cadena de nodos a lo largo de las flechas de conexión de nodo específicas 513 que conducen a un resultado positivo, denominado nodo de hoja 512. Cada nodo de hoja 512 contiene, o bien una referencia directa de documento que tiene una lista de referencias a documentos coincidentes o a parte de un documento de referencia como los metadatos, a una referencia de índice o simplemente a otros datos. De hecho, esto es cierto para cada nodo 502 que puede ser el último de una cadena de nodos con la combinación de palabras clave respectiva.

La figura 6 muestra un árbol 500 del sistema de almacenamiento y de búsqueda de documentos 200 o 300, y describe cómo se procesa la información durante un proceso de búsqueda. El usuario 400 (de la figura 3) se autenticó a sí mismo en el sistema de almacenamiento y de búsqueda de documentos, desbloqueó su material clave 410 y, por lo tanto, tiene acceso a los índices 430 que pertenecen a los documentos en los que el usuario tiene acceso al documento. Este acceso puede restringirse a su lectura, también puede ser un acceso para modificar o eliminar el documento, o para agregar más usuarios. A continuación, el usuario proporciona un término de búsqueda 550. Un término de búsqueda 550 comprende una cantidad de palabras clave o, más en general, una cantidad de posibles entradas de datos en el índice accesible 311 del índice de almacenamiento 309. El término "accesible" se refiere a los derechos de pertenencia/acceso al grupo mostrado a modo de ejemplo en la figura 4. El término de búsqueda 550 también puede comprender parte de una palabra, una fórmula o una imagen como entrada de datos; comprende elementos de formato definidos de acuerdo con las reglas de la base de datos relacionadas con las posibles entradas de datos.

De acuerdo con la invención, se aplica la primera palabra clave en el índice 430 que busca un primer nodo de árbol que se descifra. No existe necesariamente una relación 1:1 entre "palabra clave" y "nodo", los nodos también pueden contener partes (por ejemplo, el primer carácter o partes de la palabra clave) de una palabra clave. Además, pueden contener no solo un carácter, parte de una palabra o una palabra completa, sino más bien una recopilación de caracteres o palabras. A continuación, el motor de búsqueda descifra un nodo 502 de la siguiente capa o un nodo vecino, mientras que los otros tres nodos del árbol en el mismo nivel permanecen sin descifrar. Este procedimiento se repite para cada capa, en el que solo se descifra un nodo hasta que se alcanza el nodo de hoja 512. Ya que la información relevante de documento se almacena junto con un nodo, en este caso el nodo hoja 512, el resultado de búsqueda 555 se transmite de vuelta al usuario que realiza la consulta. Como se ha mencionado anteriormente, el resultado de la búsqueda puede ser una o más referencias de documentos, así como metadatos, datos semánticos relevantes o una referencia a otras fuentes de datos.

La figura 7 muestra un terminal de usuario 210. Este puede ser un ordenador con una CPU 1, una memoria 2, un almacenamiento 3 y un dispositivo de almacenamiento auxiliar 4. Por lo general, se integra una interfaz de red 5 y se proporciona una interfaz de entrada 6 y una interfaz de salida 7. La interfaz de salida 7 está conectada con una pantalla 10, mientras que la interfaz de entrada 6 está conectada con un teclado 8 y/o un ratón de ordenador 9. El terminal de usuario dispara el iniciador 201 de la figura 1 que siempre es un servidor de aplicaciones 202. El usuario inicia una búsqueda a través del terminal de usuario 210, que a continuación se reenvía al servidor de aplicaciones 202 y al sistema de búsqueda.

La figura 8A y la figura 8B muestran un diagrama de flujo de un registro en el sistema de almacenamiento y de

búsqueda de acuerdo con la figura 1; La figura 9 muestra un diagrama de flujo del sistema de almacenamiento y de búsqueda de acuerdo con la figura 1.

El procedimiento de registro de un documento se inicia con la autenticación de un usuario dispuesto a entrar en un nuevo documento en el almacenamiento de archivos 208 y en la base de datos 206.

Un cliente autenticado, que se asigna a "un usuario" en el sistema, envía una solicitud de registro de un documento 309 al servidor de aplicaciones 202 (que es, para el procedimiento de indexación posterior, el "iniciador"). El servidor de aplicaciones 202 genera una clave de documento 305 y cifra el documento con la clave de documento 305. El servidor de aplicaciones 202 registra la información de documento y las claves de documento 305 en la base de datos de documentos 206; 304 y almacena la información de documento como el documento 309 en el almacenamiento de archivos 208; 308. Finalmente, el servidor de aplicaciones 202 almacena un testigo para descifrar el documento 309 y el índice asociado o relacionado con este documento (por razones de simplicidad, en este caso no se agrega ninguna referencia en la figura 3 y en la figura 4) que a continuación se usan para procedimiento de indexación posterior.

El servidor de aplicaciones envía una solicitud para registrar el documento al servidor distribuidor 204. Esto sucede de manera asíncrona, es decir, independiente del procedimiento de almacenamiento de documentos. El servidor distribuidor 204 prepara el documento a registrar. Esto puede implicar cargar información adicional con respecto al documento y/o al índice, así como preprocesar el propio documento. Además de eso, el servidor distribuidor 204 concede acceso al material clave relevante al servidor de búsqueda 203. Esto es necesario para indexar el documento más adelante en el proceso. En otras palabras, el servidor de búsqueda 203 necesita descifrar tanto el índice a actualizar como el documento (almacenado, cifrado) a indexar. Una forma de lograr esto (pero no limitado a) es almacenando una copia de la clave de índice, así como la clave de documento, encriptadas con una clave de aplicación que sea accesible para el servidor de búsqueda. El servidor distribuidor 204 carga la clave de documento almacenada temporalmente 305 y descifra el documento 309 a indexar. El servidor distribuidor prepara el documento a indexar y lo reenvía al servidor de búsqueda 203. El servidor de búsqueda 203 carga la clave de índice almacenada temporalmente 307 desde la base de datos de índices 306. La clave de índice almacenada 307 se usa para descifrar los nodos del índice asociado que necesita actualizarse. El servidor de búsqueda 203 extrae texto del documento 309, ejecuta un análisis de lenguaje que se refiere a un diccionario que incluye palabras clave y términos semánticos, y obtiene términos y posiciones de los términos. Además, el servidor de búsqueda 203 puede obtener e incluir una pluralidad de términos, términos anteriores y/o sucesivos, ejecutar un análisis de lenguaje específico basándose en el lenguaje reconocido del documento 309. Además, el servidor de búsqueda 203 puede obtener adicionalmente información para determinar la clasificación búsqueda-resultado, tal como las frecuencias de los términos y otra información que puede extraerse y registrarse usando el procesamiento relacionado con el lenguaje del documento.

A continuación, el servidor de búsqueda actualiza la información de índice diferente 311, encripta la información de índice 311 de este documento en relación con las diferentes claves de índice 307 y la registra en el almacenamiento de índices 310. Una vez que el índice está actualizado, el acceso temporal a la clave de índice se descarta (es decir, se abandona el testigo). El servidor de búsqueda envía el resultado del procesamiento al distribuidor y finaliza la manipulación del documento recién registrado. Se prefiere que se use un índice invertido para la preparación de la base de datos de índices.

La recuperación de documentos registrados se explicará en conexión con la divulgación de la figura 9. El usuario cliente autenticado envía una consulta de búsqueda que comprende los términos de búsqueda 550 al servidor de búsqueda 203. El servidor de búsqueda carga la clave de índice 307 desde la base de datos de índices 306. La clave de índice cargada 307 es la clave de índice permitida para todos los documentos disponibles para este usuario.

El servidor de búsqueda extrae todos los términos de la consulta, es decir, la consulta de búsqueda, que comprende un número de palabras clave, se separa en una lista de palabras clave diferentes, que están dispuestas de una manera predeterminada, como de costumbre en la manipulación de consultas de búsqueda.

A continuación, el servidor de búsqueda descifra un subconjunto/primer nodo del índice 311 usando la clave de índice 307. El servidor de búsqueda verifica si toda la búsqueda ya ha revelado un resultado positivo (114). Si no, el servidor de búsqueda vuelve a la etapa de descifrado 115 para descifrar el siguiente nodo. De lo contrario, el servidor de búsqueda evalúa si el contexto de búsqueda incluye otros índices. En caso afirmativo, el servidor de búsqueda vuelve a la etapa de descifrado inicial 115. De lo contrario, el servidor de búsqueda devuelve los resultados de búsqueda 117 al cliente y finaliza la consulta.

**Lista de señales de referencia**

1 CPU
2 memoria
3 almacenamiento
305 clave de documento
306 base de datos de índices
307 clave de índice

4 dispositivo de almacenamiento auxiliar
5 interfaz de red
6 interfaz de entrada
7 interfaz de salida
8 teclado
9 ratón de ordenador
10 pantalla
200 sistema de almacenamiento y de búsqueda
201 iniciador
202 servidor de aplicaciones
203 servidor de búsqueda
204 servidor distribuidor
205 base de datos de usuarios y claves
206 base de datos de documentos
207 base de datos de índices
208 almacenamiento de archivos
209 almacenamiento de índices
210 terminal de usuario
300 sistema de almacenamiento de documentos y de búsqueda
301 terminal de usuario
302 base de datos de usuarios y claves
303 clave de usuario
304 Base de datos de documentos
308 almacenamiento de archivos
309 documento
310 almacenamiento de índices
311 índice
400 usuario
401 usuario A
402 usuario B
410 material de clave
411 material de clave del usuario A
412 material de clave del usuario B
421 testigo del usuario A
422 testigo del usuario B
430 índices
435 índices de grupo
440 entrada de autorización asociada
445 usuario miembro de grupos
500 árbol de decisión
502 nodo de árbol
503 flecha de conexión de nodo
512 nodo de hoja
513 flecha de conexión de nodo
550 término de búsqueda
555 referencia de documento de resultados

**REIVINDICACIONES**

1. Un sistema de búsqueda (200, 300) que comprende:

- un dispositivo de almacenamiento (205, 206, 207, 208, 209; 302, 304, 306, 308, 310); y
- un servidor de búsqueda (203) configurado para buscar datos almacenados en el dispositivo de almacenamiento,

en donde el dispositivo de almacenamiento (205, 206, 207, 208, 209; 302, 304, 306, 308, 310) comprende:

- un almacenamiento de archivos (208; 308) configurado para almacenar al menos un documento (309),
- un almacenamiento de índices (209; 310) configurado para almacenar al menos un índice (311), cada uno de estos asociado a al menos uno del al menos un documento (309),
- una base de datos de claves de documento (206; 304) configurada para almacenar al menos una clave de documento (305) asociada a el al menos un documento (309) en el almacenamiento de archivos (208; 308) y en donde la clave de documento (305) se usa para cifrar y descifrar el documento asociado (309),
- una base de datos de índices (207; 306) configurada para almacenar al menos una clave de índice (307), asociada a al menos un índice (311) en el almacenamiento de índices (209; 310) y en donde la clave de índice (307) se usa para cifrar y descifrar el índice asociado (311) y
- una base de datos de claves de usuario (206, 302) configurada para almacenar al menos una clave de usuario (303),

en donde cada clave de usuario (303) está asociada a una pluralidad de claves de índice (307) y a una pluralidad de claves de documento (305), usando la clave de usuario (303) respectiva para cifrar y descifrar las claves de índice asociadas (307) en la base de datos de índices (207; 306) así como las claves de documento asociadas (305) en la base de datos de claves de documento (206; 304),
en donde el servidor de búsqueda (203) está configurado para extraer uno o más términos para una búsqueda de una consulta de búsqueda solicitada, para descifrar la clave de índice cifrada (307) con la clave de usuario (303) que pertenece al usuario (301) que solicita la consulta de búsqueda para buscar en el almacenamiento de índices (209; 310), basándose en las claves de índice descifradas (307), la aparición del uno o más términos de la consulta de búsqueda solicitada, y enviar un resultado de búsqueda a un terminal, en el donde dicho índice (311) está cifrado y comprende al menos un árbol de índices (500) que comprende una secuencia de nodos interconectados (502, 512), cada uno cifrado individualmente y que comprende una referencia a otro índice, así como una información (503, 513) relevante para los nodos secundarios interconectados.

2. El sistema de búsqueda de acuerdo con la reivindicación 1, en el que cada nodo (502, 512) comprende una o más referencias en la base de datos de claves de documento (206; 304), en donde el resultado de la búsqueda comprende unos punteros a documentos coincidentes (309) en el almacenamiento de archivos (208; 308), a otros índices y/o para buscar metadatos tales como partes del índice.

3. El sistema de búsqueda de acuerdo con la reivindicación 1 o la reivindicación 2, que comprende una estructura de índices que permite un descifrado basado en nodos y niveles, descifrando solo por lo tanto las partes relevantes del índice con el fin de ejecutar una búsqueda, en el que cada nodo de árbol (502) contiene información de índice y referencias a otros nodos, nodos secundarios y nodos de hoja (512) que contienen referencias al o a los documentos coincidentes, otros índices y metadatos de búsqueda.

4. El sistema de búsqueda de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en el que la búsqueda del almacenamiento de índices (209; 310) comprende un descifrado jerárquico de nodos.

5. El sistema de búsqueda de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en el que la búsqueda comprende una búsqueda dentro de múltiples secuencias de nodo y/o múltiples árboles en la base de datos de índices.

6. El sistema de búsqueda de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que cada nodo está configurado para comprender una estructura de datos del grupo que abarca una matriz, una lista y un hashmap.

7. El sistema de búsqueda de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que la estructura de datos de cada nodo está configurada para comprender elementos del grupo que abarca una palabra clave, partes de una palabra, un carácter, una pluralidad de caracteres o palabras, entradas formateadas e imágenes.

8. El sistema de acuerdo con la reivindicación 1, que comprende un servidor de aplicaciones (202) para agregar documentos (309) a una base de datos de documentos cifrados existente (208, 308), estando el servidor de aplicaciones (202) configurado para enviar un documento al servidor de búsquedas (203), que está configurado para extraer una pluralidad de términos clave del documento (309) que se solicita que se grabe, para cargar al menos una clave de índice (307) de la base de datos de índices (306), para descifrar los índices asociados (311) de las claves de índice (307) para identificar los índices (311) a actualizar, para procesar información del documento para

actualizar los índices (311), para cifrar el índice actualizado (311) así como las claves de índice asociadas (307) y para almacenarlos en el almacenamiento de índices (310) así como en la base de datos de índices (306) respectivamente.

9. El sistema de acuerdo con la reivindicación 8, en el que se verifica la información del documento para actualizar los índices (311) para identificar los nodos de árbol de los índices relevantes (311) a actualizar, para actualizar los nodos de árbol incluyendo la creación de nuevos nodos de árbol (502) y/o nuevas conexiones (503) a otros nodos (502).

10. El sistema de acuerdo con las reivindicaciones 8 o 9, en el que el servidor de aplicaciones (202) cifra el documento (309) a registrar con la clave de documento (305), cifra la clave de documento (305) con la clave de usuario (303) que pertenece al usuario que solicita registrar el documento (309) y registra el documento cifrado en el almacenamiento de documentos (309) y la clave de documento cifrado en la base de datos de documentos (308).

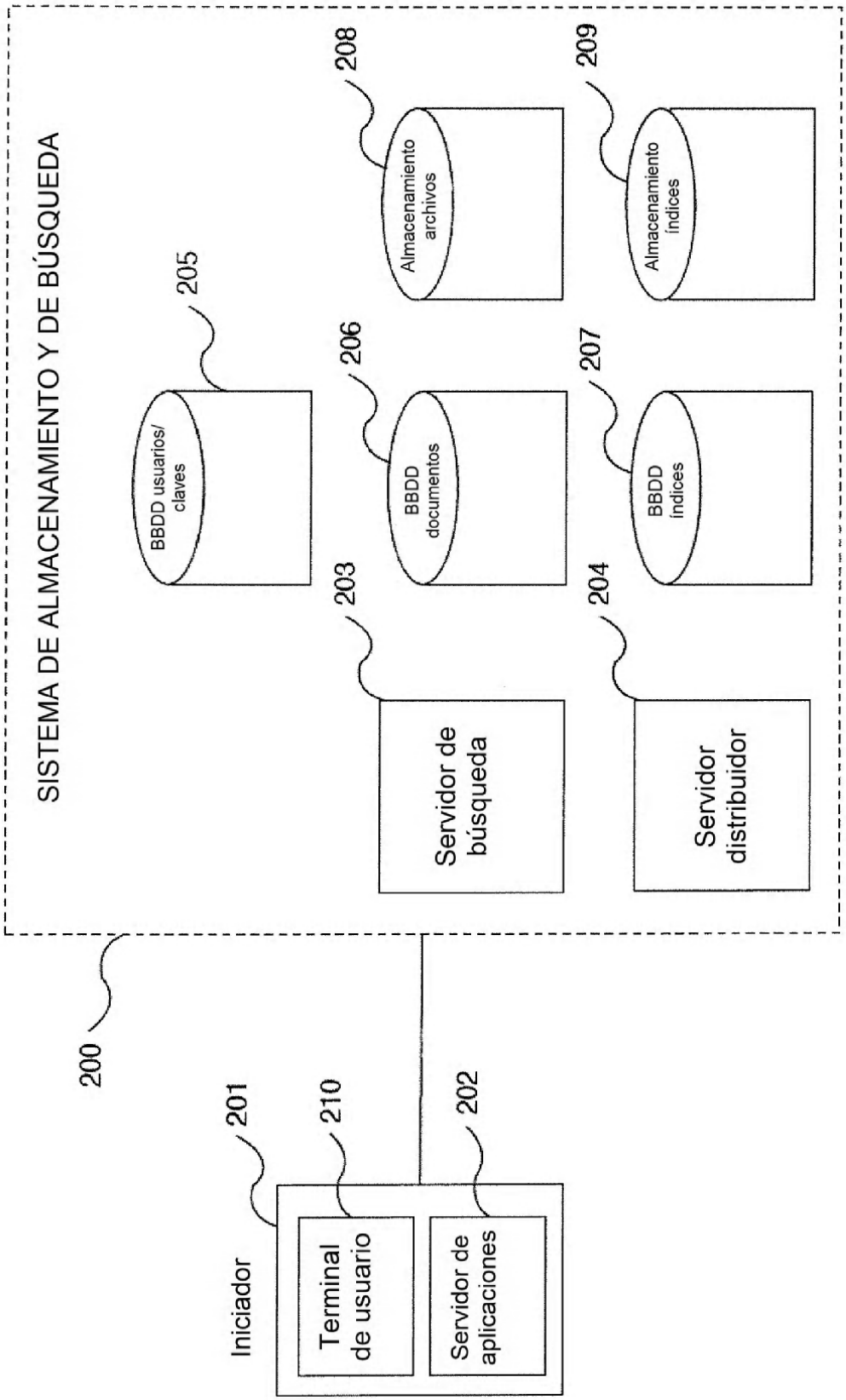
SISTEMA DE ALMACENAMIENTO Y DE BÚSQUEDA
200
205
BBDD usuarios/ claves
206
BBDD documentos
207
BBDD índices
208
Almacenamiento archivos
209
Almacenamiento índices
203
Servidor de búsqueda
204
Servidor distribuidor
201
Iniciador
210
Terminal de usuario
202
Servidor de aplicaciones

Fig. 1

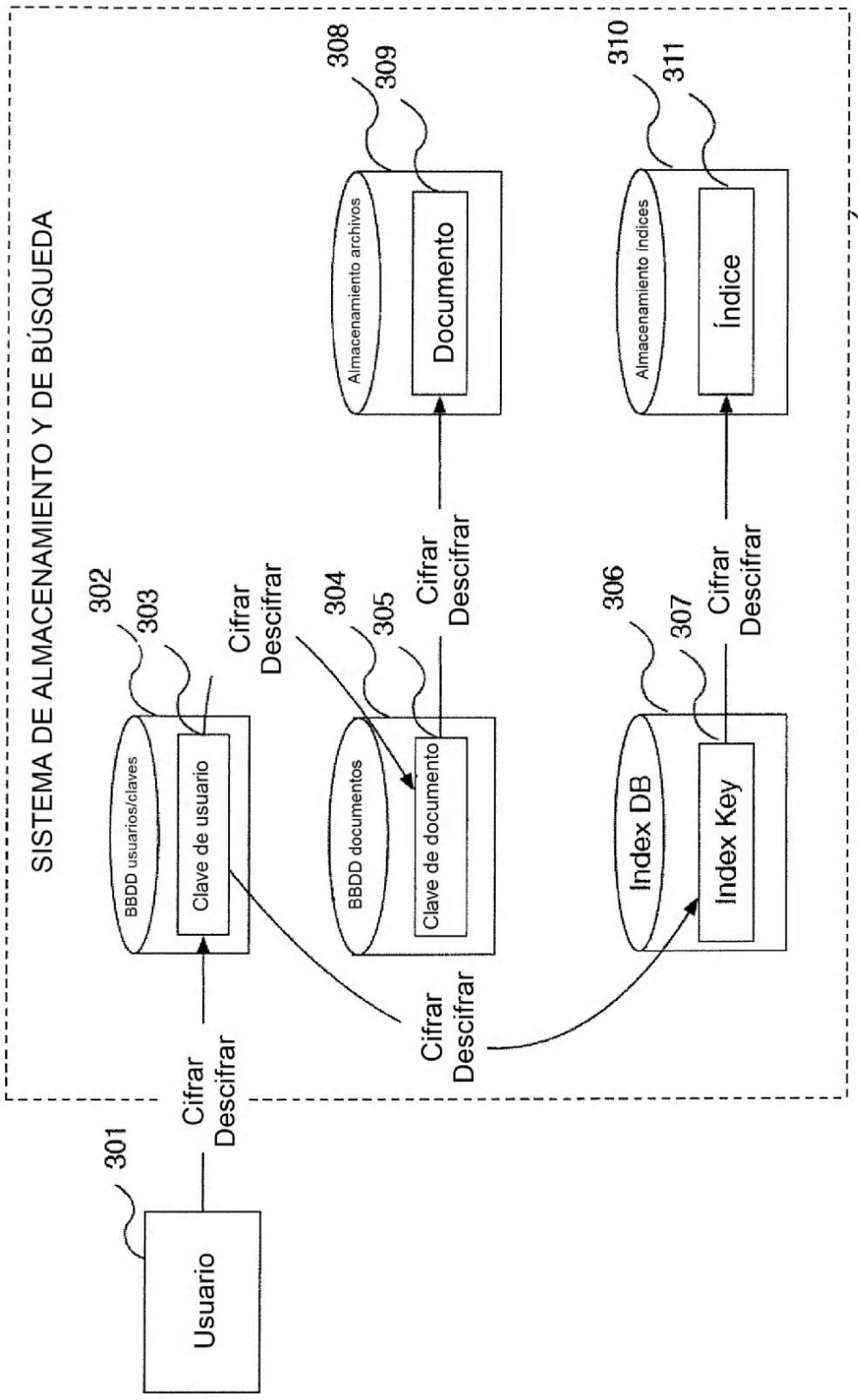
SISTEMA DE ALMACENAMIENTO Y DE BÚSQUEDA
300
Usuario
301
Cifrar
Descifrar
BBDD usuarios/claves
302
Clave de usuario
303
Cifrar
Descifrar
BBDD documentos
304
Clave de documento
305
Cifrar
Descifrar
Index DB
306
Index Key
307
Cifrar
Descifrar
Almacenamiento archivos
308
Documento
309
Cifrar
Descifrar
Almacenamiento índices
310
Índice
311
Cifrar
Descifrar

Fig. 2

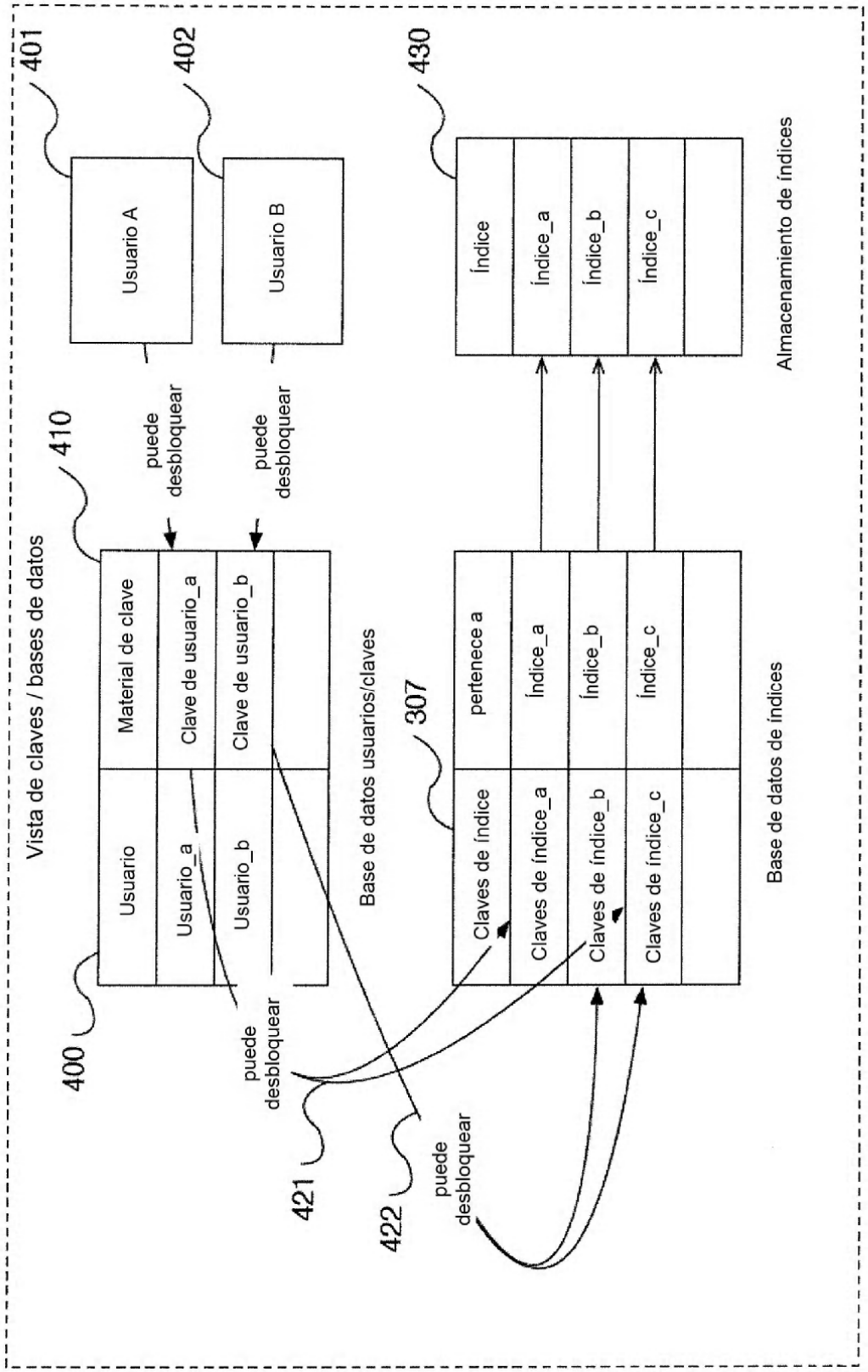
401
402
430
Usuario A
Usuario B
Índice
Índice_a
Índice_b
Índice_c
Almacenamiento de índices
410
puede desbloquear
puede desbloquear
Vista de claves / bases de datos
Material de clave
Clave de usuario_a
Clave de usuario_b
Usuario
Usuario_a
Usuario_b
Base de datos usuarios/claves
307
pertenece a
Índice_a
Índice_b
Índice_c
Claves de índice
Claves de índice_a
Claves de índice_b
Claves de índice_c
Base de datos de índices
400
puede desbloquear
421
422
puede desbloquear

Fig. 3

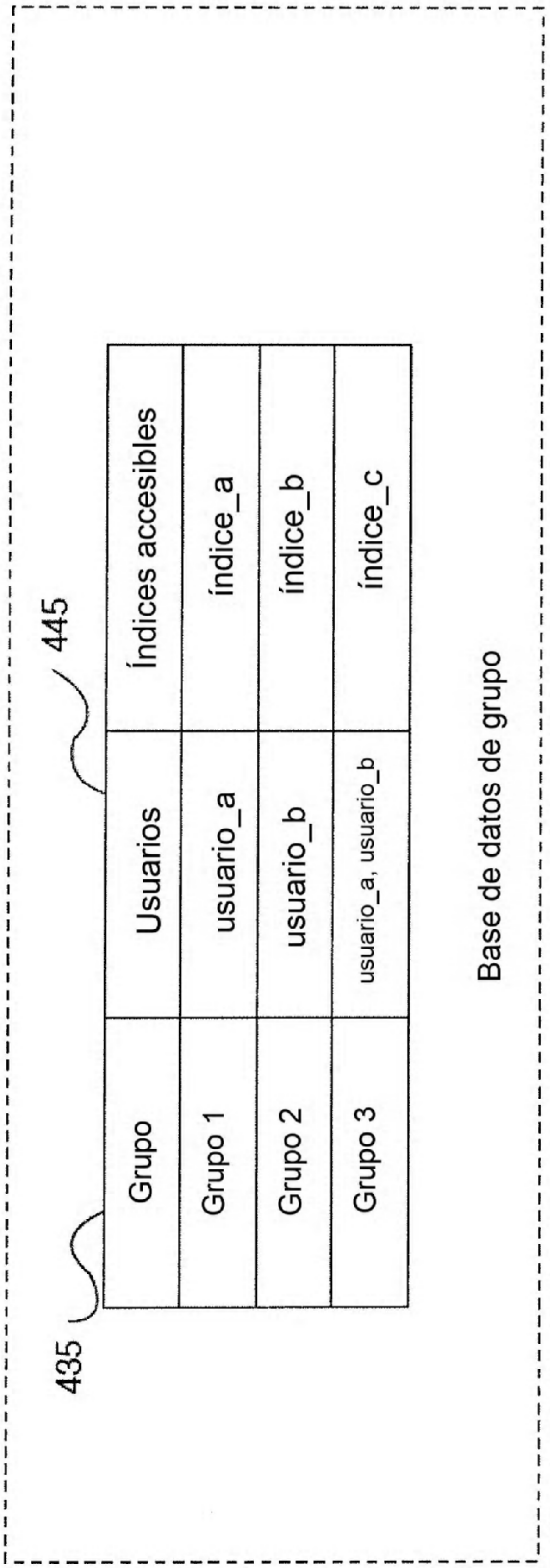
Base de datos de grupo
435
445
Grupo
Usuarios
Índices accesibles
Grupo 1
usuario_a
índice_a
Grupo 2
usuario_b
índice_b
Grupo 3
usuario_a, usuario_b
índice_c

Fig. 4

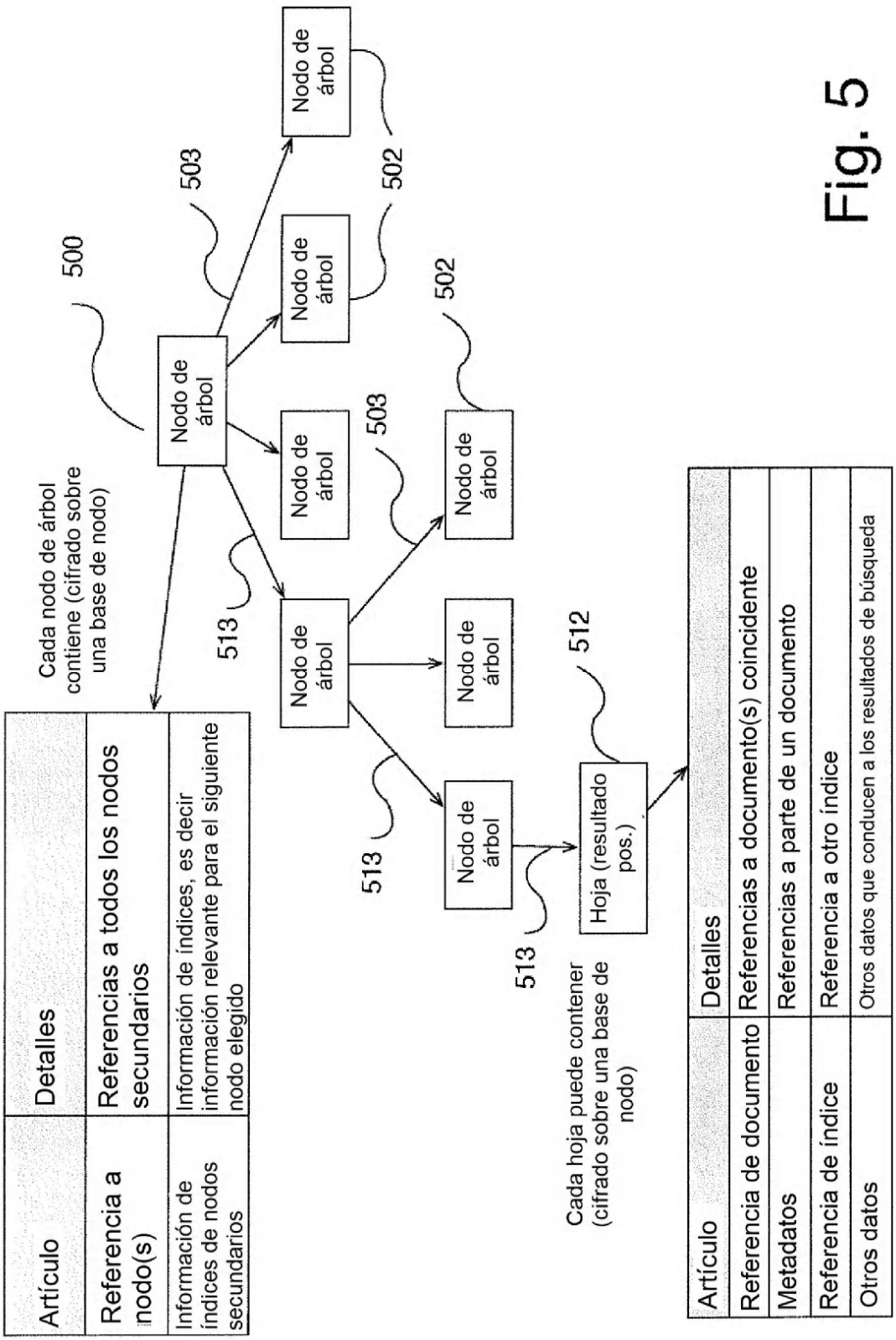
Artículo
Detalles
Referencia a nodo(s)
Referencias a todos los nodos secundarios
Información de índices de nodos secundarios
Información de índices, es decir información relevante para el siguiente nodo elegido
Cada nodo de árbol contiene (cifrado sobre una base de nodo)
500
Nodo de árbol
503
502
513
512
Hoja (resultado pos.)
Cada hoja puede contener (cifrado sobre una base de nodo)
Artículo
Detalles
Referencia de documento
Referencias a documento(s) coincidente
Metadatos
Referencias a parte de un documento
Referencia de índice
Referencia a otro índice
Otros datos
Otros datos que conducen a los resultados de búsqueda

Fig. 5

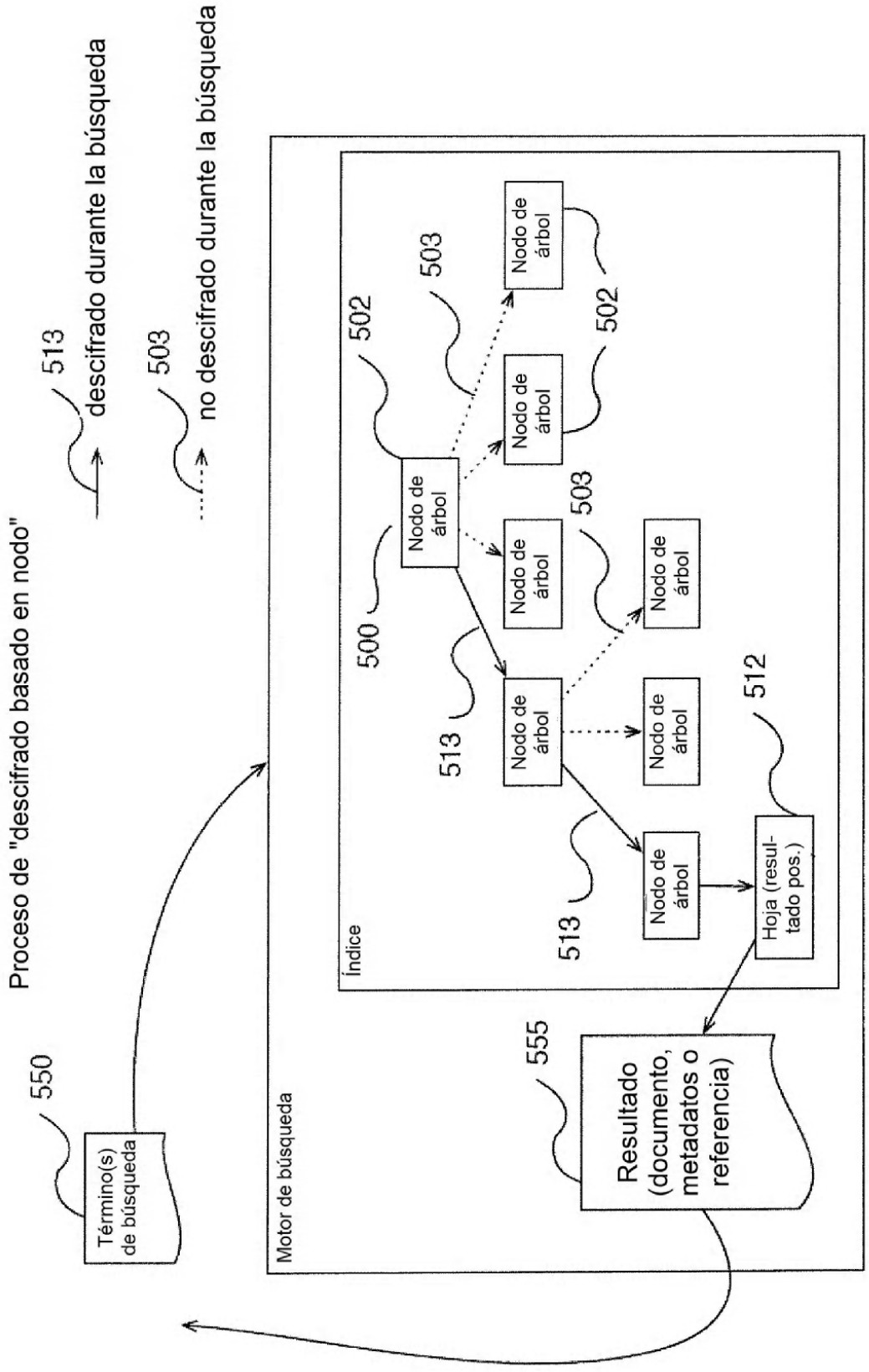
Proceso de "descifrado basado en nodo"
513
descifrado durante la búsqueda
503
no descifrado durante la búsqueda
550
Término(s) de búsqueda
Motor de búsqueda
Índice
500
502
503
513
512
Nodo de árbol
Hoja (resultado pos.)
555
Resultado (documento, metadatos o referencia)

Fig. 6

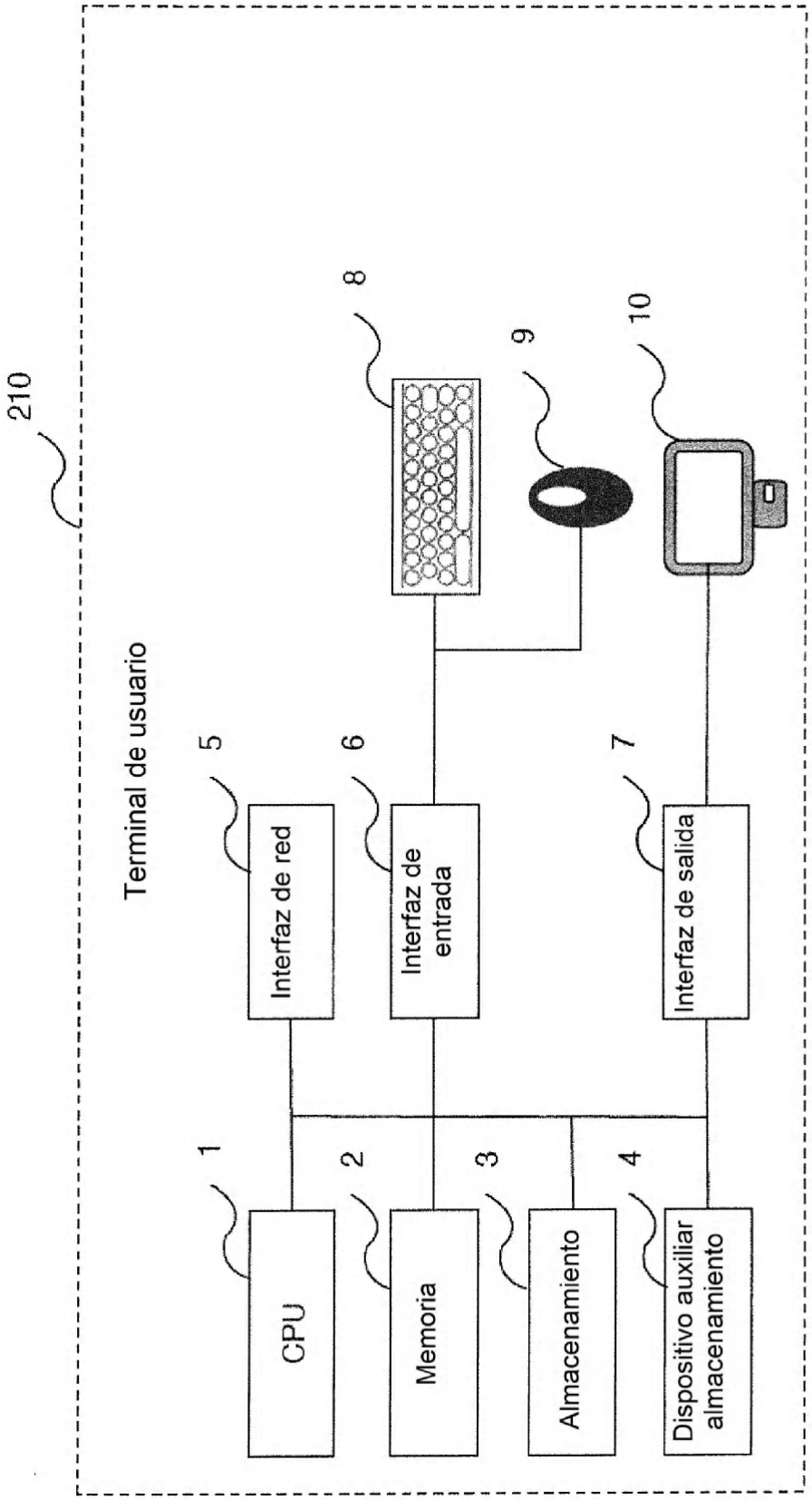
210
Terminal de usuario
1
CPU
2
Memoria
3
Almacenamiento
4
Dispositivo auxiliar almacenamiento
5
Interfaz de red
6
Interfaz de entrada
7
Interfaz de salida
8
9
10

Fig. 7

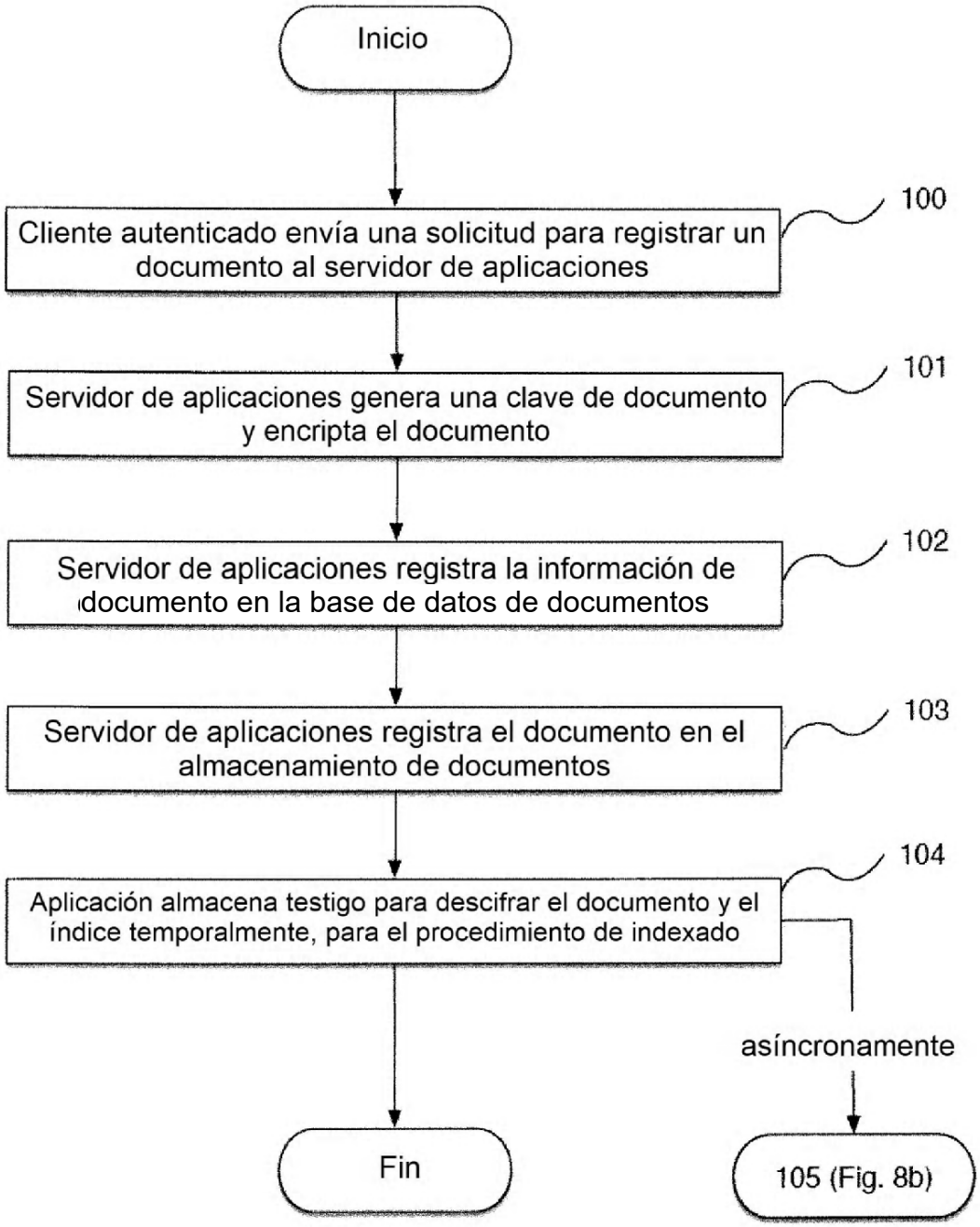
Inicio
Cliente autenticado envía una solicitud para registrar un documento al servidor de aplicaciones
100
Servidor de aplicaciones genera una clave de documento y encripta el documento
101
Servidor de aplicaciones registra la información de documento en la base de datos de documentos
102
Servidor de aplicaciones registra el documento en el almacenamiento de documentos
103
Aplicación almacena testigo para descifrar el documento y el índice temporalmente, para el procedimiento de indexado
104
asíncronamente
Fin
105 (Fig. 8b)

Fig. 8a

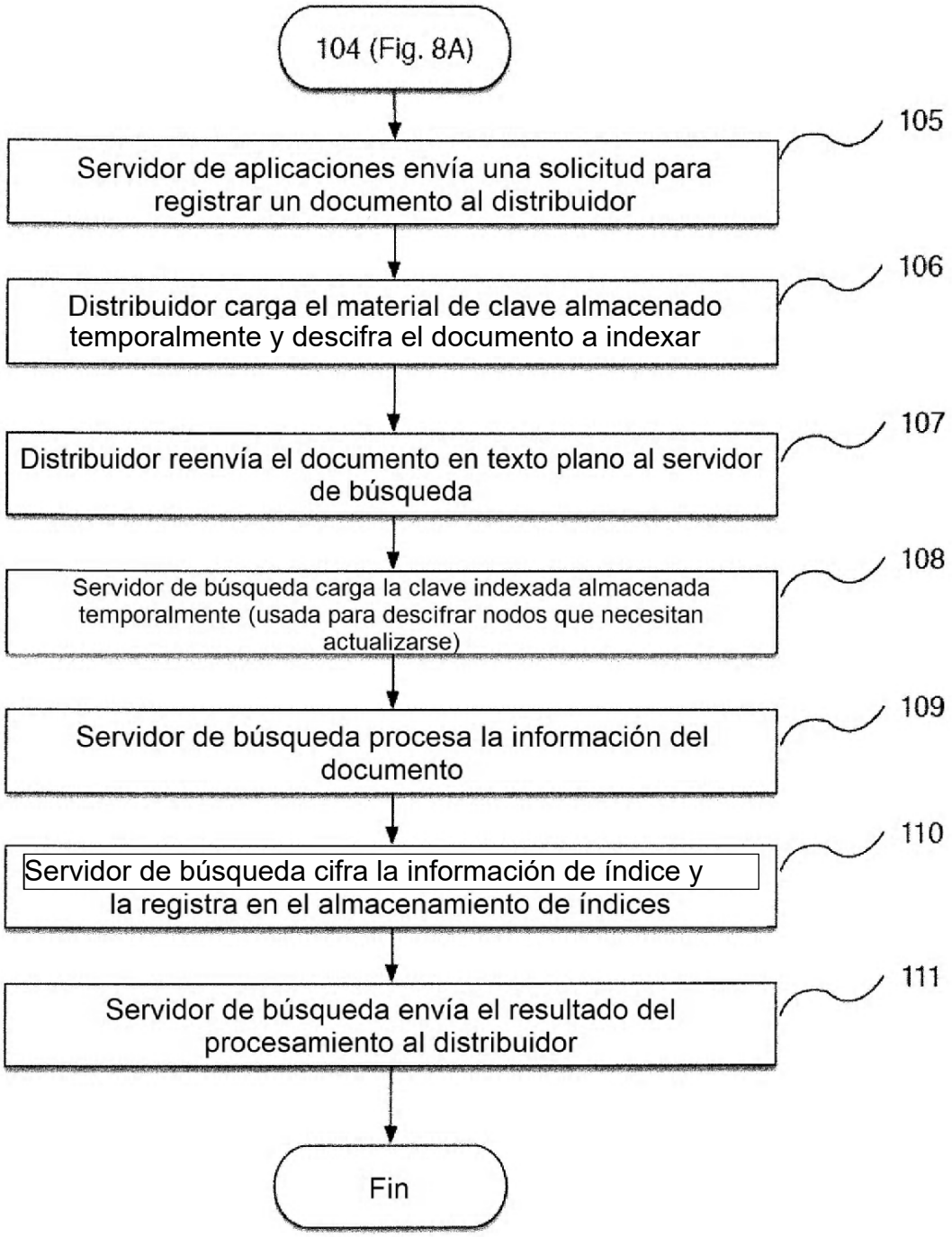
104 (Fig. 8A)
105
Servidor de aplicaciones envía una solicitud para registrar un documento al distribuidor
106
Distribuidor carga el material de clave almacenado temporalmente y descifra el documento a indexar
107
Distribuidor reenvía el documento en texto plano al servidor de búsqueda
108
Servidor de búsqueda carga la clave indexada almacenada temporalmente (usada para descifrar nodos que necesitan actualizarse)
109
Servidor de búsqueda procesa la información del documento
110
Servidor de búsqueda cifra la información de índice y la registra en el almacenamiento de índices
111
Servidor de búsqueda envía el resultado del procesamiento al distribuidor
Fin

Fig. 8b

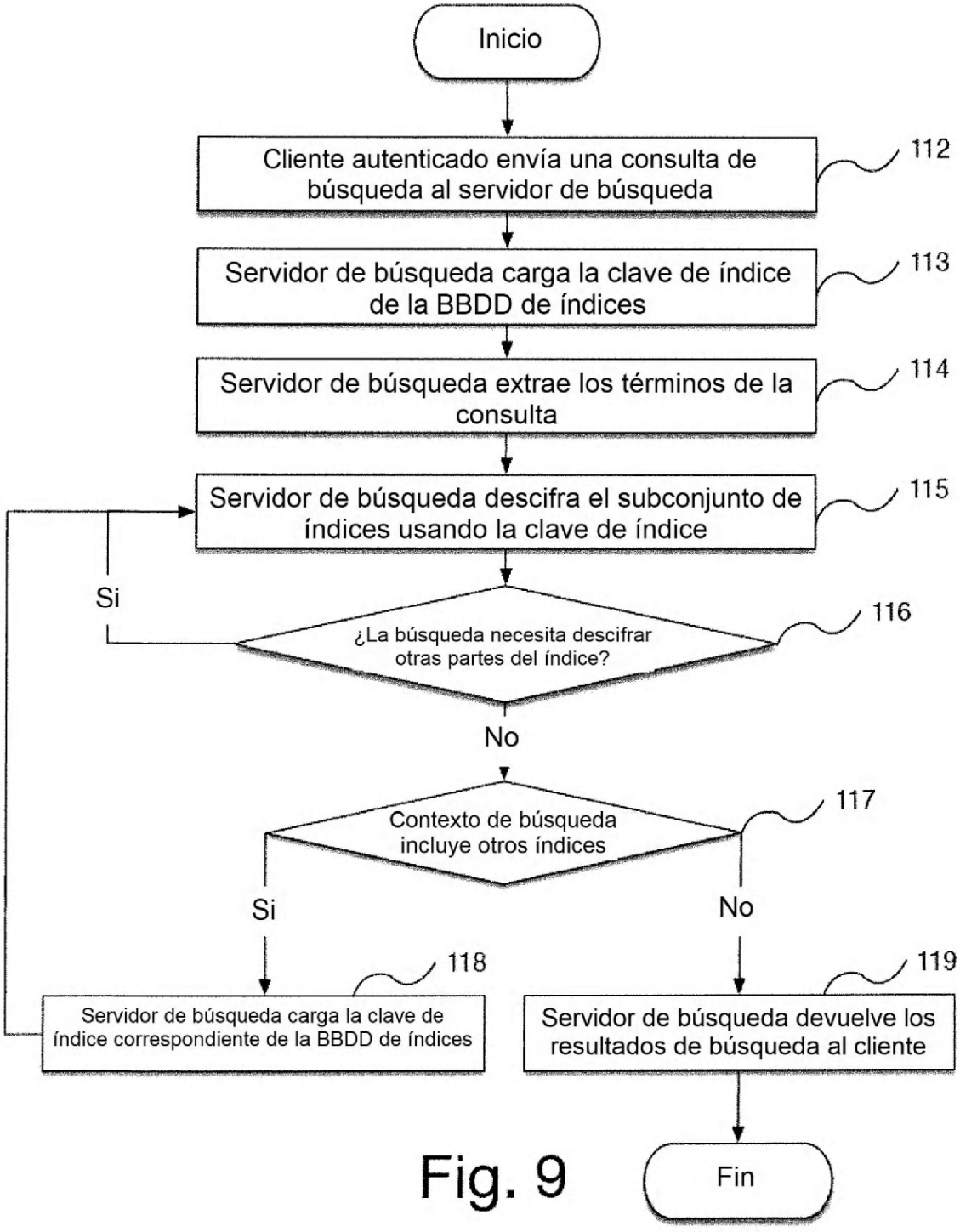
Inicio
Cliente autenticado envía una consulta de búsqueda al servidor de búsqueda
112
Servidor de búsqueda carga la clave de índice de la BBDD de índices
113
Servidor de búsqueda extrae los términos de la consulta
114
Servidor de búsqueda descifra el subconjunto de índices usando la clave de índice
115
Si
¿La búsqueda necesita descifrar otras partes del índice?
116
No
Contexto de búsqueda incluye otros índices
117
Si
No
Servidor de búsqueda carga la clave de índice correspondiente de la BBDD de índices
118
Servidor de búsqueda devuelve los resultados de búsqueda al cliente
119
Fin

Fig. 9