

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 855 074**

51 Int. Cl.:

G06F 16/2455 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.04.2019** **E 19166538 (9)**

97 Fecha y número de publicación de la concesión europea: **02.12.2020** **EP 3550447**

54 Título: **Realización de la adaptación de actualización de caché**

30 Prioridad:

03.04.2018 FR 1852865
03.04.2018 US 201815944657

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
23.09.2021

73 Titular/es:

AMADEUS S.A.S. (100.0%)
485 route du Pin Montard, Sophia Antipolis
06410 Biot, FR

72 Inventor/es:

CANIS, LAURE;
BERTRAND, JEROME;
HERER, MAREK;
RONDEPIERRE, THOMAS;
MURTADAK, DIVENDAR UMESH;
PASQUIER-MEUNIER, NICOLAS;
MORETTI, REMI y
SAUCH, FRANCIS

74 Agente/Representante:

SUGRAÑES, S.L.P.

ES 2 855 074 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

Realización de la adaptación de actualización de caché

5 Campo técnico

La divulgación objeto se refiere en general a ordenadores y software informático, y en particular a métodos, sistemas y un producto de programa informático que manejan consultas de búsqueda en un sistema de base de datos y realizan la adaptación de actualización de caché.

10 Antecedentes

Los desarrollos recientes en la tecnología de base de datos muestran que es un problema común asegurar tiempos de respuesta cortos para buscar consultas que requieren procesar grandes volúmenes de datos. Por ejemplo, tal procesamiento ha de realizarse en respuesta a unas denominadas "consultas abiertas" que contienen únicamente poca información de entrada (por ejemplo se especifica únicamente uno o dos parámetros de una docena de posibles parámetros y/o los intervalos de valor especificados de los parámetros son amplios) y, en consecuencia, conduce a un gran número de resultados en general. Las posibilidades para acelerar el procesamiento de datos aumentando el rendimiento de hardware son limitadas. Por lo tanto, se llama la atención a mejorar los mecanismos que subyacen al procesamiento de grandes volúmenes de datos.

Un enfoque general para acortar tiempos de respuesta a consultas es calcular previamente o recopilar previamente resultados para consultas de búsqueda y mantenerlas en una caché. Las consultas de búsqueda realmente entonces no se procesan en los grandes volúmenes de datos, sino en los resultados según se mantienen en la caché.

El almacenamiento en caché, sin embargo, tiene una desventaja, en concreto que los resultados mantenidos en la caché pueden quedar desactualizados debido a cambios en los datos a partir de los que se han calculado previamente o recopilado previamente los resultados y, por lo tanto, difieren de los datos. Es por lo tanto un problema, mantener los resultados calculados previamente o recopilados previamente actualizados para asegurar que las consultas devueltas por los resultados de la caché reflejen correctamente los correspondientes datos entre volúmenes de datos. Son por lo tanto necesarias estrategias para mantener la caché actualizada, es decir, los resultados mantenidos en la caché.

Diversas estrategias relativamente sencillas para mantener la caché actualizada, como son conocidas en la técnica anterior, se refieren, por ejemplo, a recalcular o volver a recopilar todos los resultados mantenidos en la caché de manera frecuente, establecer y mantener planificaciones de recálculo o recopilación manualmente y recalcular o volver a recopilar los resultados cuando se consideran desactualizados. En general, recalcular o volver a recopilar un gran número de resultados de búsqueda (o actualizar una caché) es una tarea que consume una cantidad sustancial de recursos computacionales así como recursos de ancho de banda. Por lo tanto, se buscan técnicas de recálculo o de recopilación para realizar eficazmente el recálculo o recopilación de los resultados de búsqueda.

Se analiza otra estrategia para acortar los tiempos de respuesta mientras se asegura la validez de los resultados de búsqueda con cierta probabilidad en el documento EP 2 908 255 A1. De acuerdo con esta estrategia, se estima la validez de resultados de búsqueda calculados previamente o recopilados previamente y esta estimación de validez se utiliza para devolver resultados de búsqueda calculados previamente o recopilados previamente que probablemente son válidos al cliente. Más específicamente, se proporciona un método de manejo de consultas en un sistema de base de datos que tiene al menos un cliente y al menos una plataforma de búsqueda. La plataforma de búsqueda mantiene resultados de búsqueda calculados previamente que están asociados con factores de confianza. Un factor de confianza indica una probabilidad de que el resultado de búsqueda asociado sea válido. Se recibe una consulta que indica al menos un criterio de búsqueda por la plataforma de búsqueda. La plataforma de búsqueda utiliza los factores de confianza asociados con los resultados de búsqueda recopilados previamente identificados para aumentar la probabilidad media de que los resultados de búsqueda recopilados previamente que son válidos se devuelvan al cliente.

55 Sumario

De acuerdo con un primer aspecto de la presente divulgación, se proporciona un método de procesamiento de consultas en un sistema de base de datos. El sistema de base de datos comprende una fuente de datos original que almacena una pluralidad de piezas de datos originales y una fuente de caché que almacena una pluralidad de piezas de datos almacenados en caché, en donde cada pieza de datos almacenados en caché está asociada con un valor de precisión. El método comprende seleccionar aleatoriamente consultas de entre una pluralidad de consultas manejadas por el sistema de base de datos, en el momento en el que se recibe la respectiva consulta desde un cliente. Además, el método comprende, para cada una de las consultas seleccionadas aleatoriamente, recuperar una primera pieza de los datos almacenados en caché que coincide con la consulta seleccionada aleatoriamente desde la fuente de caché, y recuperar una primera pieza de los datos originales que coincide con la consulta seleccionada aleatoriamente desde la fuente de datos original. También, el método comprende, para consultas que no se

seleccionan aleatoriamente, recuperar una segunda pieza de los datos almacenados en caché que coincide con la consulta desde la fuente de caché, evaluar el valor de precisión de la segunda pieza de los datos almacenados en caché, y si el valor de precisión se encuentra por debajo de un umbral dado, recuperar una segunda pieza de los datos originales que coincide con la consulta desde la fuente de datos original, y actualizar la segunda pieza de los datos almacenados en caché por la segunda pieza de los datos originales. El método comprende la adaptación de un modelo probabilístico basándose en la primera pieza de los datos almacenados en caché y la primera pieza de los datos originales usando un algoritmo de aprendizaje de máquina. El valor de precisión se deriva del modelo probabilístico e indica una probabilidad de que una pieza de los datos almacenados en caché coincida con una correspondiente pieza de los datos originales.

De acuerdo con un ejemplo, el método comprende, para las consultas seleccionadas aleatoriamente, comparar la primera pieza de los datos almacenados en caché y la primera pieza de los datos originales, actualizar la primera pieza de los datos almacenados en caché mediante la primera pieza de los datos originales, y devolver la primera pieza de los datos originales al cliente.

En otro ejemplo, el método comprende adicionalmente, para consultas que no se seleccionan aleatoriamente, devolver la segunda pieza de los datos originales como resultado al cliente si el valor de precisión se encuentra por debajo del umbral dado, y de lo contrario, devolver la segunda pieza de los datos almacenados en caché como resultado al cliente.

En otro ejemplo más, el método comprende adicionalmente, para cada una de las consultas seleccionadas aleatoriamente, enviar información acerca de la consulta, la primera pieza de los datos almacenados en caché y la primera pieza de los datos originales a una base de datos de historial del sistema de base de datos.

De acuerdo con un ejemplo adicional, el método comprende adicionalmente comparar la primera pieza de los datos originales con la primera pieza de los datos almacenados en caché, almacenar un primer valor indicador que indica si la primera pieza de los datos originales coincide con la primera pieza de los datos almacenados en caché, comparar el valor de precisión de la primera pieza de los datos almacenados en caché y el umbral dado, almacenar un segundo valor indicador que indica si el valor de precisión se encuentra por debajo del umbral dado, y actualizar el umbral dado basándose en pares del primer y el segundo valores indicadores.

En el ejemplo adicional, actualizar el umbral dado puede comprender opcionalmente determinar una tasa de interrogación y una eficacia de interrogación basándose en los pares del primer y segundo valores indicadores, y actualizar el umbral dado basándose en la tasa de interrogación y la eficacia de interrogación.

Opcionalmente además, en el ejemplo adicional, la tasa de interrogación puede determinarse basándose en la ecuación: $(TP+FP) / (TP+FP+TN+FN)$, y la eficacia de interrogación puede determinarse basándose en la ecuación: $TP / (TP+FP)$. En las ecuaciones, TP es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales no coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión se encuentra por debajo del umbral dado. TN es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión no se encuentra por debajo del umbral dado. FP es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión se encuentra por debajo del umbral dado. FN es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales no coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión no se encuentra por debajo del umbral dado.

De acuerdo con un segundo aspecto de la divulgación objeto, se proporciona un sistema de base de datos de procesamiento de consultas. El sistema de base de datos comprende una fuente de datos original que almacena una pluralidad de piezas de datos originales y una fuente de caché que almacena una pluralidad de piezas de datos almacenados en caché, estando asociada cada pieza de datos almacenados en caché con un valor de precisión. El sistema de base de datos está configurado adicionalmente para seleccionar aleatoriamente consultas de entre una pluralidad de consultas manejadas por el sistema de base de datos, en el momento que se recibe la respectiva consulta desde un cliente. También, el sistema de base de datos está configurado para, para cada una de las consultas seleccionadas aleatoriamente, recuperar una primera pieza de los datos almacenados en caché que coincide con la consulta seleccionada aleatoriamente desde la fuente de caché, y recuperar una primera pieza de los datos originales que coincide con la consulta seleccionada aleatoriamente desde la fuente de datos original. El sistema de base de datos está configurado para, para consultas que no se seleccionan aleatoriamente, recuperar una segunda pieza de los datos almacenados en caché que coincide con la consulta desde la fuente de caché, evaluar el valor de precisión de la segunda pieza de los datos almacenados en caché, y si el valor de precisión se encuentra por debajo de un umbral dado, recuperar una segunda pieza de los datos originales que coincide con la consulta desde la fuente de datos original, y actualizar la segunda pieza de los datos almacenados en caché mediante la segunda pieza de los datos originales. El sistema de base de datos está configurado adicionalmente para adaptar un modelo probabilístico basándose en la primera pieza de los datos almacenados en caché y la primera pieza de los datos originales usando un algoritmo de aprendizaje de máquina. El valor de precisión se deriva del modelo probabilístico e indica una

probabilidad de que una pieza de los datos almacenados en caché coincida con una correspondiente pieza de los datos originales.

De acuerdo con un tercer aspecto de la divulgación objeto, se proporciona un producto de programa informático. El producto de programa informático comprende instrucciones que, cuando se ejecutan por un ordenador, hacen que el ordenador realice el método de acuerdo con los aspectos de la primera realización.

El sumario anterior presenta un sumario simplificado para proporcionar un entendimiento básico de algunos aspectos de los sistemas y/o métodos analizados en el presente documento. Este sumario no es una vista general extensiva de los sistemas y/o métodos analizados en el presente documento. No se pretende para identificar elementos clave/críticos o para delinear el alcance de tales sistemas y/o métodos. Su único propósito es presentar algunos conceptos de una forma simplificada como un preludio a una descripción más detallada que se presenta más adelante.

Breve descripción de los dibujos

Los dibujos adjuntos ilustran diversas realizaciones de la divulgación objeto y, junto con la descripción general proporcionada anteriormente, y la descripción detallada de las realizaciones proporcionadas a continuación, sirven para explicar las realizaciones de la divulgación objeto. En los dibujos, se usan números de referencia similares para indicar partes similares en las diversas vistas.

La Figura 1 muestra un sistema de base de datos de ejemplo, en el que puede realizarse un método de manejo de consultas de búsqueda.

La Figura 2 muestra un método de ejemplo de manejo de consultas de búsqueda en el sistema de base de datos de ejemplo de la Figura 1.

La Figura 3 muestra los principios de la divulgación objeto para procesar consultas y adaptar la decisión de interrogación.

La Figura 4 muestra principios adicionales de la divulgación objeto para procesar consultas y adaptar la decisión de interrogación.

La Figura 5 muestra un método de procesamiento de consultas de acuerdo con la divulgación objeto.

La Figura 6 muestra un sistema de base de datos de ejemplo, en el que puede realizarse el método mostrado en la Figura 5.

La Figura 7 muestra un método de ejemplo para el manejo de consultas de búsqueda no seleccionadas en el método de la Figura 5.

La Figura 8 muestra un método de ejemplo de la actualización de datos almacenados en caché en el método de la Figura 5.

La Figura 9 muestra un método de ejemplo de la actualización un umbral en el sistema de base de datos de la Figura 6.

La Figura 10 muestra una representación esquemática de un sistema informático de acuerdo con una realización.

Descripción detallada

Antes de volver a la descripción detallada de las realizaciones, se explicarán en primer lugar algunos aspectos más generales sobre las técnicas implicadas.

La divulgación objeto en general pertenece al manejo de consultas de búsqueda en un sistema de base de datos. La expresión "consulta de búsqueda" incluye todos los tipos de solicitudes de base de datos que incluyen, por ejemplo, solicitudes de lectura para recuperar datos y escribir solicitudes para insertar, cambiar o borrar datos. Se muestra un sistema de base de datos 1 de ejemplo para manejar consultas de búsqueda en la Figura 1.

El sistema de base de datos 100 incluye al menos una plataforma de búsqueda 110 y al menos una base de datos 120. Para aumentar la seguridad frente a fallos o el rendimiento, puede estar presente una pluralidad de plataformas de búsqueda 110. Puede haber también una pluralidad de bases de datos 120 para aumentar la capacidad o el rendimiento. La pluralidad de bases de datos 120 pueden estar ubicadas en un sitio, o pueden estar distribuidas a través de una pluralidad de sitios. La al menos una plataforma de búsqueda 110 y la al menos una base de datos 120 están conectados a través de una red tal como una Red de Área Local (LAN), una Red de Área Metropolitana (MAN), o una Red de Área Extensa (WAN) tal como Internet.

La al menos una base de datos 120 almacena datos originales. En el presente documento, tales bases de datos 120 que almacenan los datos originales se denominan también como una fuente de datos original. En más detalle, la al menos una base de datos 120 almacena una pluralidad de piezas de datos originales. En este punto, la expresión "datos originales" significa datos que son válidos y que forman la base para todo el procesamiento de datos en el sistema de base de datos 100, en particular los datos a los que se dirigen las consultas de búsqueda. La expresión "base de datos" en este punto se pretende que abarque cualesquiera tipos de sistema de almacenamiento de información estructurado tales como bases de datos independientes convencionales como bases de datos de SQL o no de SQL así como bases de datos complejas, distribuidas y/o propietarias, bases de datos relacionales o bases de datos orientadas a objetos. Finalmente, el término "pieza" en este punto significa cualquier tipo de información

estructurada tal como registros, series, partes de tablas de base de datos (por ejemplo, columnas, líneas, o entradas), o similares.

Las piezas de datos originales pueden referirse a cualquier tipo de información en un dominio particular. Por ejemplo, los datos originales pueden estar ubicados en el dominio de las simulaciones informáticas de circuitos electrónicos (tales como simulaciones de circuitos sujetos a la decisión T 1227/05 por la Comisión de Apelación de la EPO) y contienen información sobre valores de entrada para las simulaciones informáticas de circuitos electrónicos, parámetros de un modelo usado por las simulaciones informáticas de circuitos electrónicos, resultados de simulación intermedios y/o finales y similares. Estos datos relacionados con la simulación informática de circuitos electrónicos se mantienen en la al menos una base de datos 120 y pueden consultarse por un sistema informático en el que se lleva a cabo la simulación informática de circuitos electrónicos, por ejemplo, mediante la al menos una plataforma de búsqueda 110, para recuperar valores o parámetros de entrada, escribir o actualizar los resultados de simulación.

Además, en el sistema de base de datos 100, la al menos una plataforma de búsqueda 110 mantiene resultados de búsqueda calculados previamente o recopilados previamente para reducir los tiempos de respuesta para contestar consultas de búsqueda en comparación con tiempos requeridos para recuperar los correspondientes datos originales desde la al menos una base de datos 120 y realizar el cálculo en los mismos para obtener los resultados de búsqueda. En este sentido, se usan las expresiones "calculados previamente" y "recopilados previamente" para cubrir cualquier clase de cálculo previo y recopilación previa tal como rastreadores de Internet que recopilan o copian el contenido de servidores web de Internet, pero también cálculos complejos e intensivos en el tiempo de resultados de búsqueda basados en datos originales subyacentes. En el dominio de las simulaciones informáticas de circuitos electrónicos, el cálculo previo puede implicar la realización de simulaciones con un conjunto de parámetros reducido (en comparación con todos los parámetros del sistema del mundo real), pero con valores variables de dicho conjunto de parámetros para obtener resultados que van a usarse como valores de entrada en las simulaciones adicionales con conjuntos adicionales de parámetros u otros.

En la al menos una plataforma de búsqueda 110, puede proporcionarse al menos una fuente de caché 115 para almacenar los resultados de búsqueda. Los datos almacenados en la al menos una fuente de caché 115 se denominan como datos almacenados en caché y puede almacenarse una pluralidad de piezas de datos almacenados en caché en la al menos una fuente de caché 115. Por ejemplo, las piezas de datos almacenados en caché almacenadas en la al menos una fuente de caché 115 pueden corresponder a piezas de datos originales en la al menos una base de datos 120 (es decir, la fuente de datos original) y la pluralidad de las piezas de datos almacenados en caché en la al menos una fuente de caché 115 se refieren a un subconjunto de la pluralidad de piezas de datos originales desde la al menos una base de datos 120.

Sin embargo, como ya se ha mencionado anteriormente, almacenando en caché resultados de búsqueda calculados previamente o recopilados previamente mantenidos en la al menos una fuente de caché 115 de la al menos una plataforma de búsqueda 110 quedan desactualizados con el tiempo debido a cambios en los datos originales desde los que se han calculado previamente o recopilado previamente. Debido a los cambios en los datos originales, ocurren diferencias entre los correspondientes datos almacenados en caché y los datos originales y, por lo tanto, los datos almacenados en caché quedan inválidos.

Para proporcionar una indicación de la probabilidad de si los datos almacenados en caché son válidos o no, los resultados de búsqueda recopilados previamente o calculados previamente mantenidos en la al menos una fuente de caché 115 mediante la plataforma de búsqueda 110 están asociados con valores de precisión (también denominados como factores de confianza en el presente documento). Los valores de precisión pueden ser similares a los factores de confianza como se describe por el documento EP 2 908 255 A1. Por ejemplo, cada resultado de búsqueda recopilado previamente o calculado previamente almacenado por la plataforma de búsqueda 110 tiene un correspondiente valor de precisión. Como alternativa, un valor de precisión puede estar asociado con una pluralidad de resultados de búsqueda recopilados previamente o calculados previamente. Un valor de precisión indica una probabilidad de que el resultado o resultados de búsqueda recopilados previamente o calculados previamente sean válidos. Es decir, el valor de precisión indica una probabilidad de que una pieza de los datos almacenados en caché almacenada en la al menos una fuente de caché 115 coincida con una correspondiente pieza de datos originales, es decir, una pieza de datos originales desde la que se ha derivado la pieza de datos almacenados en caché. El valor de precisión asociado con los resultados de búsqueda recopilados previamente o calculados previamente se utiliza para decidir si devolver los resultados de búsqueda recopilados previamente o calculados previamente desde la al menos una fuente de caché 115 o desde la al menos una base de datos 120.

El modelo probabilístico, desde el que se derivan en general los valores de precisión de los resultados de búsqueda calculados previamente o recopilados previamente en la fuente de caché 115, modela las discrepancias entre los resultados de búsqueda almacenados en caché (calculados previamente o recopilados previamente) y los resultados de búsqueda (es decir, asumidos o si ya se han recuperado realmente) basándose en los datos originales en la al menos una base de datos 120. Es decir, el modelo probabilístico aproxima la precisión o imprecisión de cualquier resultado de búsqueda almacenado en caché. Como un ejemplo, el modelo probabilístico modela la validez probable de los resultados de búsqueda almacenados en caché a lo largo del tiempo.

Pueden emplearse diversos tipos de modelos probabilísticos. Por ejemplo, el modelo probabilístico puede ser un modelo conceptual que se ha generado por un proceso de modelado estructurado conocido como modelado científico. El modelado requiere identificar y seleccionar aquellos aspectos y eventos del entorno del mundo real (en este punto, por ejemplo, los aspectos y eventos que tienen influencia sobre la volatilidad y precisión que se reduce a lo largo del tiempo de los resultados de búsqueda almacenados en caché) y dejando de lado otros aspectos que pueden ser de ninguna relevancia particular para el fin del modelo. Adicionalmente, los aspectos identificados y seleccionados pueden abstraerse y formularse/implementarse en forma procesable tal como un programa informático y/o ecuaciones matemáticas. Un modelo conceptual de este tipo puede servir también como un medio para simulación informática que simula el comportamiento (es decir, la volatilidad/precisión) de los resultados de búsqueda calculados previamente o recopilados previamente a lo largo del tiempo. Como alternativa, el modelo probabilístico empleado puede ser un modelo histórico-estadístico que está monitorizando y evaluando, por ejemplo, la volatilidad/precisión de los resultados de búsqueda calculados previamente o recopilados previamente en el periodo de tiempo dado en el pasado (por ejemplo, tres meses) y extrapola las tendencias determinadas en el futuro.

Además de modelar la precisión usando el modelo probabilístico, se evita una caída grave de la precisión siendo reactivo a eventos. Las decisiones de recálculo/recopilación se definen en la recepción de eventos predeterminados que pueden tener un impacto sobre la corrección de los resultados de búsqueda almacenados en caché. Los eventos pueden ser asíncronos, es decir, el punto de tiempo de su ocurrencia no está predeterminado - pueden ocurrir en cualquier momento, o ser síncronos, es decir, el punto de tiempo de su ocurrencia puede determinarse con antelación - pueden ocurrir en una base regular.

En la Figura 1, al menos un cliente 130 (normalmente una pluralidad de clientes) está conectado a la al menos una plataforma de búsqueda 110 a través de una red tal como una Red de Área Local (LAN), Red de Área Metropolitana (MAN), Red de Área Extensa (WAN) tal como Internet, o una red de comunicación móvil tal como la Evolución a Largo Plazo (LTE). La conexión puede ser alámbrica o inalámbrica. Por ejemplo, los clientes 130 pueden ser ordenadores personales o dispositivos de comunicación móvil tal como teléfonos inteligentes, tabletas, asistencia personal o similares.

A continuación, se explicará cómo el sistema de base de datos 100 mostrado en la Figura 1 maneja consultas de búsqueda.

La Figura 2 ilustra un diagrama de flujo de un método 200 para manejar consultas de búsqueda en el sistema de base de datos 100 de la Figura 1.

En la operación 210, la pluralidad de clientes 130 dirigen consultas de búsqueda a la plataforma de búsqueda 110 que incluyen uno o más criterios o parámetros de búsqueda. Por ejemplo, si una consulta de búsqueda es una búsqueda de Internet, la consulta de búsqueda puede llevar a cabo una cadena de búsqueda, texto de búsqueda o frase de búsqueda como criterios de búsqueda. Un criterio de búsqueda adicional puede ser el idioma de los sitios web en los que va a buscarse o una indicación de un punto de tiempo de la primera disponibilidad de la cadena de búsqueda solicitada, texto de búsqueda o frase de búsqueda. En el ejemplo anterior, en el dominio de las simulaciones informáticas, la consulta de búsqueda puede llevar, como criterios de búsqueda, el nombre de un parámetro de entrada para el que es necesario un valor por la simulación informática.

La plataforma de búsqueda 110 recibe las consultas de búsqueda, procesa cada una de las consultas de búsqueda recibidas desde los clientes 130 y realiza correspondientes búsquedas en los resultados de búsqueda recopilados previamente o calculados previamente. En más detalle, en la operación 220, la plataforma de búsqueda 110 puede enviar una solicitud de búsqueda para realizar una búsqueda para cada consulta de búsqueda a la fuente de caché 115. La solicitud de búsqueda incluye al menos parte del uno o más criterios o parámetros de búsqueda de la consulta de búsqueda. La fuente de caché 115 realiza la búsqueda basándose en la solicitud de búsqueda, es decir, busca datos entre los datos almacenados en caché que coinciden con los criterios o parámetros de la solicitud de búsqueda, en la operación 225, y envía datos almacenados en caché hallados como resultados de búsqueda calculados previamente o recopilados previamente a la plataforma de búsqueda 110 en la operación 230. Los resultados de búsqueda calculados previamente o recopilados previamente (es decir, datos almacenados en caché) tienen valores o parámetros que coinciden (es decir, están adaptados) con los criterios o parámetros de búsqueda de la solicitud de búsqueda.

Cualquier resultado de búsqueda calculado previamente o recopilado previamente coincidente en la fuente de caché 115 se devuelve por la plataforma de búsqueda 110 al cliente 130 desde el cual se recibió la consulta de búsqueda como respuesta a la consulta de búsqueda en la operación 280. El cliente 130 recibe esta respuesta y presenta los resultados de búsqueda al usuario, o, en el dominio de las simulaciones informáticas, usa los mismos como entradas para la simulación llevada a cabo en las mismas.

En caso de que no se hallen resultados de búsqueda calculados previamente o recopilados previamente coincidentes en la fuente de caché 115, la plataforma de búsqueda 110 puede hacer los criterios o parámetros de búsqueda de la consulta de búsqueda menos restrictivos y realiza una búsqueda revisada enviando solicitudes de búsqueda revisadas con los criterios o parámetros de búsqueda menos restrictivos a la fuente de caché 115. La plataforma de búsqueda

110 puede realizar también la búsqueda revisada no aplicando todos los criterios o parámetros de búsqueda dados. Los detalles particulares de cuáles y cómo los criterios o parámetros de búsqueda de las consultas de búsqueda se hacen menos restrictivos, y/o cuáles criterios o parámetros de búsqueda son menos importantes en las consultas de búsqueda, pueden establecerse en la plataforma de búsqueda 110 mediante reglas dadas o determinadas usando técnicas de aprendizaje de máquina, por ejemplo, modelos probabilísticos basándose en criterios de búsqueda históricos o parámetros y resultados de búsqueda calculados previamente o recopilados previamente.

Como se ha descrito anteriormente, los resultados de búsqueda en la fuente de caché 115 pueden estar asociados con valores de precisión. En caso afirmativo, en la operación 240, la plataforma de búsqueda 110 puede evaluar los valores de precisión de los resultados de búsqueda recopilados previamente o calculados previamente que coinciden con los criterios o parámetros de búsqueda antes de devolverlos a los clientes 130, de manera similar como en el documento EP 2 908 255 A1. Por ejemplo, la plataforma de búsqueda 110 compara los valores de precisión con un umbral (dado en la consulta de búsqueda por los clientes 130, o prescrito en la plataforma de búsqueda 110). De esta manera, la plataforma de búsqueda 110 determina la probabilidad de que los resultados de búsqueda sean válidos, es decir, la probabilidad de que los resultados de búsqueda enviados por la fuente de caché 115 coincidan con los correspondientes datos originales en la base de datos 120 desde los que se han recopilado previamente o calculado previamente.

Si los valores de precisión se encuentran por encima el umbral dado, la plataforma de búsqueda 110 devuelve los resultados de búsqueda asociados con los mismos en la operación 280. De lo contrario, para los resultados de búsqueda enviados por la fuente de caché 115 con valores de precisión no por encima del umbral dado, la plataforma de búsqueda 110 recupera correspondientes datos originales desde las bases de datos 120. Es decir, la plataforma de búsqueda 110 envía las respectivas solicitudes de búsqueda a las bases de datos 120 en la operación 250, y las bases de datos 120 buscan datos originales que coinciden con los criterios o parámetros de búsqueda en las solicitudes de búsqueda, en la operación 255. A continuación, en la operación 260, las bases de datos 120 devuelven los datos originales hallados. En respuesta a la recepción de los datos originales desde las bases de datos 120, la plataforma de búsqueda 110 devuelve los datos originales recibidos como resultados de búsqueda a los clientes 130 en la operación 280. La plataforma de búsqueda 110 puede actualizar también los resultados de búsqueda recuperados de las bases de datos 120 en la fuente de caché 115 enviando los mismos, en la operación 270, a la fuente de caché 115 para su actualización. La fuente de caché 115 actualiza los datos almacenados en caché usando los resultados de búsqueda recibidos de la plataforma de búsqueda 110.

Es decir, de manera similar al documento EP 2 908 255 A1, los resultados de búsqueda calculados previamente asociados con valores de precisión que indican una probabilidad de que el resultado de búsqueda asociado sea válido, se mantienen en la caché. Sin embargo, como en el documento EP 2 908 255 A1, los valores de precisión son de una naturaleza estática de manera que la interrogación se realizará cada vez que los valores de precisión asociados con resultados de búsqueda particulares indiquen la invalidez de los mismos. Esto conduce a una gran cantidad de interrogación innecesaria, y por lo tanto ineficaz, debido al hecho de que los resultados de búsqueda recuperados desde la base de datos por medio de interrogación aún coinciden con los resultados de búsqueda calculados previamente. Es decir, los resultados de búsqueda calculados previamente son realmente aún válidos incluso si los valores de precisión asociados indican su invalidez. Además, como ya se ha mencionado anteriormente, el recálculo y la recopilación de los resultados de búsqueda es una tarea que consume una cantidad sustancial de recursos de cálculo así como recursos de ancho de banda y la divulgación objeto consigue mejoras de eficacia al actualizar los datos almacenados en caché de una manera síncrona a medida que se recuperan datos originales desde las bases de datos 120 preferentemente únicamente para piezas de datos almacenados en caché que es probable que estén desactualizados.

En el presente documento, recuperar datos originales desde la base de datos 120 y usar los mismos para actualizar datos almacenados en caché en la caché 115 o para responder a consultas de búsqueda también se denomina como interrogación u operación de interrogación. La interrogación puede realizarse en respuesta a la recepción de una consulta de búsqueda particular, es decir, de una manera síncrona o independiente de cualquier consulta de búsqueda. En el último caso, se actualiza al menos una parte, o la totalidad, de los datos almacenados en caché en la caché 115 por medio de una operación en lote, de una manera asíncrona. La decisión de si interrogar o no se denomina como decisión de interrogación. Por ejemplo, evaluar el valor de precisión asociado con los datos almacenados en caché contra un umbral dado, como en la operación 240 de la Figura 2, para determinar si los datos almacenados en caché es probable que sean válidos o no y, por lo tanto, si es necesario recuperar datos originales desde la base de datos, es una decisión de interrogación en el contexto de la divulgación objeto.

La divulgación objeto tiene en cuenta estos problemas y proporciona un método de adaptación del modelo probabilístico para aumentar la eficacia de las decisiones de interrogación reduciendo el número de piezas de interrogación de datos originales que tienen correspondientes datos válidos en la caché 115 en comparación con el documento EP 2 908 255 A1. En más detalle, la divulgación objeto permite evitar la interrogación innecesaria e ineficaz o al menos limitar su cantidad, por medio de la adaptación de los valores de precisión y por lo tanto la decisión de interrogar, haciéndolos de esta manera dinámicos. Esto permite limitar el número de solicitudes de búsqueda o consultas de búsqueda que se realiza contra las bases de datos 120 y por lo tanto resuelve las desventajas de, por ejemplo, el documento EP 2 908 255 A1. Los principios generales de la divulgación objeto se describirán ahora con

referencia a las Figuras 3 y 4.

En general, el método de procesamiento de consultas de búsqueda como se muestra en la Figura 3 selecciona un número de consultas de búsqueda 310 de entre todas las consultas de búsqueda 320 manejadas por un sistema de base de datos tal como el sistema de base de datos 100 mostrado en la Figura 1. En más detalle, en todo momento que se recibe una consulta de búsqueda en el sistema de base de datos 100 desde un cliente 130, se decide si procesar la consulta de búsqueda de la manera conocida como se ha desvelado anteriormente con referencia a la Figura 2, o seleccionar la consulta de búsqueda y manejar la misma como una consulta de búsqueda seleccionada 310. Las consultas de búsqueda seleccionadas forman consultas reales que se procesan para adaptar el modelo probabilístico y se procesan de una manera particular, como se describe a continuación. La decisión de seleccionar una consulta de búsqueda puede ser aleatoria, o estar basada en ciertos factores prescritos tales como, por ejemplo, factores específicos de consulta de búsqueda (por ejemplo, la consulta de búsqueda incluye criterios o parámetros de búsqueda prescritos), factores específicos de cliente (por ejemplo, la consulta de búsqueda se recibe desde clientes prescritos), u otros factores (por ejemplo, la consulta de búsqueda se recibe en tiempos o intervalos de tiempo prescritos). La consulta de búsqueda seleccionada 310 se procesa (operación 330) recuperando los datos almacenados en caché desde la fuente de caché 115 y recuperando también los datos originales desde la base de datos 120, haciendo coincidir tanto los criterios como los parámetros de búsqueda de la consulta de búsqueda seleccionada 310. Los datos originales recuperados de la base de datos 120 se devuelven al cliente 130 y pueden usarse también para actualizar los datos almacenados en caché en la fuente de caché 115. Los datos almacenados en caché recuperados y los datos originales recuperados se usan para adaptar el modelo probabilístico. En particular, se comparan los datos almacenados en caché recuperados y los datos originales recuperados para determinar si hubiera sido apropiado o no recuperar los datos originales. Es decir, si los resultados de comparación en esos datos almacenados en caché recuperados coinciden con los datos originales recuperados, recuperar los datos originales no hubiera sido conveniente ya que los datos almacenados en caché eran aún válidos; de lo contrario, hubiera sido sensato ya que los datos almacenados en caché eran inválidos. Este resultado que indica la eficacia de interrogación se almacena junto con los datos almacenados en caché recuperados y los datos originales recuperados (operación 340), y se usa para adaptar el modelo probabilístico (operación 350) a partir del cual se deriva el valor de precisión asociado con los datos almacenados en caché. La adaptación se realiza preferentemente usando un algoritmo de aprendizaje de máquina como se describirá en más detalle a continuación. Por consiguiente, se mejora el modelo probabilístico y se adapta la decisión de si interrogar (es decir, recuperar los datos originales desde la base de datos y actualizar los datos almacenados en caché) es necesaria y, por lo tanto, eficaz, aumentando de esta manera la precisión del modelo de probabilidad. Esto permite reducir el número de interrogaciones mejorando la decisión de interrogación en comparación con el documento EP 2 908 255 A1.

De acuerdo con estos principios, se selecciona un número de consultas de búsqueda entre todas las consultas de búsqueda manejadas por el sistema de base de datos para manejarse de manera diferente que las consultas de búsqueda no seleccionadas. Las consultas de búsqueda seleccionadas se manejan como se ha descrito con referencia a la Figura 3, mientras que las consultas de búsqueda no seleccionadas se manejan como se ha descrito con referencia a la Figura 2. Esto resulta para determinar si la interrogación hubiera sido apropiada y, por lo tanto, eficaz para los datos almacenados en caché y los correspondientes datos originales, ya que puede determinarse si los datos almacenados en caché ya se han quedado realmente desactualizados o no. Esto se hace en una manera en línea basándose en consultas de búsqueda reales de manera que puede evitarse el procesamiento fuera de línea adicional usando, por ejemplo, consultas de búsqueda de prueba o consultas de búsqueda históricas, mientras que se aumenta la precisión del modelo de probabilidad.

El método mostrado en la Figura 3 puede mejorarse adicionalmente teniendo en cuenta el valor de precisión asociado con los datos almacenados en caché y el valor umbral, como se ha descrito anteriormente. Como se muestra en la Figura 4, el método comprende adicionalmente un proceso de pronóstico 410 (que puede realizarse por un pronosticador como se describirá en conexión con la Figura 6 a continuación). En respuesta a la recuperación de los datos almacenados en caché que coinciden con la consulta de búsqueda, se evalúa el valor de precisión asociado con los datos almacenados en caché contra el valor umbral, similar a la operación 240 de la Figura 2. El resultado del proceso del pronóstico 410 indica que la interrogación se hubiera realizado (es decir, los datos originales se hubieran recuperado y devuelto puesto que los datos almacenados en caché es probable que estuvieran desactualizados) o no (es decir, los datos almacenados en caché es probable que fueran válidos y se hubieran devuelto). Este resultado que indica un proceso de decisión de pronóstico (pronosticador) también se almacena (operación 420) y se considera al adaptar el modelo probabilístico usando el algoritmo de aprendizaje de máquina (430). De esta manera, puede compararse el resultado que indica la eficacia de interrogación y el resultado que indica una decisión de pronosticador. Por ejemplo, puede determinarse si los datos almacenados en caché se han desactualizado realmente mientras se evalúa el valor de precisión contra el valor umbral que muestra que los datos almacenados en caché es probable que no estén desactualizados. Esto mejora adicionalmente el modelo probabilístico y es necesaria la decisión de si interrogar, aumentando adicionalmente de esta manera la precisión del modelo de probabilidad. Esto permite limitar el número de interrogaciones a un mínimo.

Ahora, se describirá en más detalle el método de procesamiento de consultas de búsqueda, como un aspecto de la divulgación objeto, con referencia a las Figuras 5 a 9.

El método se realiza en el sistema de base de datos 600 que comprende al menos una base de datos 120 como la fuente de datos original y al menos una caché 115 como la fuente de caché. La fuente de datos original 120 almacena una pluralidad de piezas de datos originales, y la fuente de caché 115 almacena una pluralidad de piezas de datos almacenados en caché. En el presente documento, debería entenderse que se recuperan las piezas de los datos almacenados en caché desde la fuente de datos original 120 y se almacenan en la fuente de caché 115 mediante una operación de interrogación, y por lo tanto representan versiones almacenadas en caché de piezas de datos originales. Sin embargo, los datos almacenados en caché no están limitados a versiones almacenadas en caché y pueden referirse a resultados de procesamiento de cálculo previo o recopilación previa basándose en los datos originales. La operación de interrogación debe entenderse como que comprende el procesamiento de cálculo previo o recopilación previa. Los datos almacenados en caché están asociados con un valor de precisión que indica una probabilidad de que los datos almacenados en caché coincidan con los datos originales. Por ejemplo, cada pieza de los datos almacenados en caché puede estar asociada con un valor de precisión de este tipo que indica la probabilidad de que dicha pieza de los datos almacenados en caché coincida con la correspondiente pieza de los datos originales, es decir, la pieza de los datos originales desde la que se derivaron. El valor de precisión se deriva de un modelo probabilístico establecido en el sistema de base de datos 600. El sistema de base de datos 600 comprende adicionalmente un modelador 610 para mantener y adaptar el modelo probabilístico así como derivar los valores de precisión desde el modelo probabilístico.

El sistema de base de datos 600 recibe, en la operación 510, una pluralidad de consultas de búsqueda (también denominadas brevemente como consultas en lo sucesivo) desde el al menos un cliente 130 conectado al sistema de base de datos 600 y maneja las consultas de búsqueda recibidas mediante una plataforma de búsqueda similar a la plataforma de búsqueda 110 de la Figura 1.

El método 500 de la Figura 5 se inicia con la operación 520 para seleccionar consultas de búsqueda de entre la pluralidad de consultas de búsqueda manejadas por el sistema de base de datos 600. En más detalle, el sistema de base de datos 600 decide, en el momento de la recepción de una consulta de búsqueda particular desde el al menos un cliente 130 en la operación 510, si ha de manejarse o no la consulta de búsqueda como una consulta de búsqueda seleccionada. Esto se hace en una base aleatoria, pero puede hacerse también en una base regular o de acuerdo con reglas dadas en otras realizaciones. Por ejemplo, la base regular puede incluir manejar cada consulta de búsqueda que ocupa el orden diez o cien como una consulta de búsqueda seleccionada. Las reglas dadas pueden definir criterios o parámetros de búsqueda que si están presentes en una consulta de búsqueda designan la consulta de búsqueda para seleccionarse, u otros factores para la designación para seleccionarse, tales como tiempo de recepción, número dado de consultas de búsqueda con criterios de búsqueda o parámetros similares o idénticos, número dado de consultas de búsqueda recibidas desde el mismo cliente 130, o modelos estadísticos. La consulta de búsqueda se maneja como la consulta de búsqueda seleccionada si se designa por las reglas dadas. En el sistema de base de datos 600, un selector 620 que recibe la consulta de búsqueda, una indicación de que se ha recibido la consulta de búsqueda, una copia de la consulta de búsqueda, o similares, lo realiza la operación 520 y designa la consulta de búsqueda como la consulta de búsqueda seleccionada.

En respuesta a que se haya seleccionado la consulta de búsqueda en la operación 520 (es decir, para la consulta de búsqueda que se ha seleccionado aleatoriamente), se recupera una pieza de datos almacenados en caché (denominada como primera pieza de datos almacenados en caché) que coincide con la consulta de búsqueda desde la fuente de caché 115 en la operación 530. Más específicamente, se proporciona la fuente de caché 115 con una solicitud de búsqueda que incluye al menos parte de los criterios o parámetros de búsqueda incluidos en la consulta de búsqueda. La fuente de caché 115 busca datos almacenados en caché entre todos los datos almacenados en caché almacenados en la misma, que coinciden con la solicitud de búsqueda, particularmente los criterios o parámetros de búsqueda de la misma. Los datos almacenados en caché que coinciden con la solicitud de búsqueda se devuelven, como la primera pieza de datos almacenados en caché, desde la fuente de caché 115. Esta operación 530 es similar a las operaciones 220, 225, y 230 mostradas en la Figura 2. Si no se hallan datos almacenados en caché que coinciden con la solicitud de búsqueda, el método 500 puede continuar adicionalmente con datos almacenados en caché que indican ninguna coincidencia.

Además, en la operación 540, se recupera una pieza de datos originales (denominada como primera pieza de datos originales) que coincide con la consulta de búsqueda desde la fuente de datos original 120. Más específicamente, se proporciona la fuente de datos original 120 con una solicitud de búsqueda que incluye al menos parte de los criterios o parámetros de búsqueda incluidos en la consulta de búsqueda. La solicitud de búsqueda enviada a la fuente de datos original 120 corresponde en vista de los criterios o parámetros de búsqueda incluidos en la solicitud de búsqueda enviada a la fuente de caché 115. La fuente de caché 115 busca datos originales que coinciden con la solicitud de búsqueda, particularmente los criterios o parámetros de búsqueda de la misma. Los datos originales que coinciden con la solicitud de búsqueda se devuelven, como la primera pieza de datos originales, desde la fuente de datos original 120. Esta operación 540 es similar a las operaciones 250, 255, y 260 mostradas en la Figura 2. En algunos ejemplos, las operaciones 530 y 540 se realizan en paralelo. Si no se hallan datos originales que coincidan con la consulta de búsqueda, el método 500 puede continuar adicionalmente con datos originales que no indican coincidencia.

En algunos ejemplos, si no se hallan datos almacenados en caché en la fuente de caché 115 que coinciden con la consulta de búsqueda en la operación 530, los criterios o parámetros de búsqueda de la consulta de búsqueda pueden

hacerse menos restrictivos, como se ha descrito anteriormente en conexión con la Figura 2 y a continuación se realiza la operación 530 para la consulta de búsqueda con los criterios o parámetros de búsqueda menos restrictivos. También, si no se hallan datos originales en la fuente de datos original 120 que coincidan con la consulta de búsqueda en la operación 540, se devuelve una correspondiente información que indica que no están presentes datos originales que coincidan con la consulta de búsqueda en la fuente de datos original 120 o se devuelven datos almacenados en caché hallados en la operación 530 al cliente 130 en la operación 570.

El método 500 puede incluir también una operación 550 para actualizar la primera pieza de datos almacenados en caché en la fuente de caché 115 usando la primera pieza de datos originales recuperados desde la fuente de datos original 120. La operación 550 es similar a la operación 270 mostrada en la Figura 2.

A continuación, en la operación 560, se adapta el modelo probabilístico basándose en la primera pieza de datos almacenados en caché y la primera pieza de datos originales. La operación 560 puede realizarse para cada consulta de búsqueda seleccionada aleatoriamente en la operación 520. En general, la operación 560 no se realiza individualmente para cada consulta de búsqueda seleccionada, sino de una manera asíncrona para un número recopilado de consultas de búsqueda seleccionadas aleatoriamente en la operación 520. Por ejemplo, la realización de manera asíncrona de la operación 520 significa realizar la operación 520 cuando se ha seleccionado y procesado un número dado de consultas de búsqueda. Especialmente en caso de realización de la operación 560 de la manera asíncrona, se recupera la primera pieza de datos almacenados en caché y la primera pieza de datos originales mediante las operaciones 530 y 540 en el sistema de base de datos 600 para su procesamiento posterior en la operación 560. La adaptación del modelo probabilístico en la operación 560 puede usar un algoritmo de aprendizaje de máquina, como se describirá en el presente documento a continuación. La operación 560 puede realizarse por el modelador 610, que también está configurado para recibir la primera pieza de datos almacenados en caché y la pieza de datos originales para cada consulta de búsqueda que se ha seleccionado aleatoriamente.

Finalmente, o en paralelo con cualquier operación 550 o 560, se devuelve la primera pieza de datos originales recuperados desde la fuente de datos original 120 al cliente 130 como el resultado de búsqueda en la operación 570. En este punto, en caso de que los criterios o parámetros de búsqueda de la consulta de búsqueda se hayan hecho más restrictivos como se ha descrito anteriormente, devolver la primera pieza de datos originales incluye una indicación de cuáles y cómo se han cambiado los criterios o parámetros de búsqueda de la consulta de búsqueda de modo que son menos restrictivos.

En un ejemplo, la operación 560 incluye comparar las primeras piezas de datos almacenados en caché y las primeras piezas de datos originales y determinar un valor estadístico que indica una probabilidad de que la primera pieza de datos almacenados en caché coincida con la primera pieza de datos originales. En este caso, es decir, si ambas de las primeras piezas coinciden, recuperar la pieza de datos originales desde la fuente de datos original 120 no hubiera sido necesario puesto que la pieza de datos almacenados en caché era aún válida.

En respuesta a la adaptación del modelo probabilístico, pueden actualizarse los valores de precisión de todas o al menos algunas piezas de datos almacenados en caché en la fuente de caché 115. Por ejemplo, los valores de precisión de los datos almacenados en caché pueden actualizarse derivándolos del modelo probabilístico adaptado.

En la divulgación objeto, puede establecerse cualquier algoritmo de aprendizaje de máquina que pueda usarse en relación con modelos probabilísticos. Por ejemplo, el algoritmo de aprendizaje de máquina, dada una pluralidad de consultas de búsqueda seleccionadas convertidas en un conjunto de características estructuradas, crea un modelo probabilístico para predecir un objetivo (por ejemplo, primer valor indicador almacenado con las consultas de búsqueda). El término "característica" en este punto se refiere a cualquier información que pueda deducirse a partir de las consultas de búsqueda. La conversión de la pluralidad de consultas de búsqueda seleccionadas puede comprender calcular valores de todas las características, usando los criterios de búsqueda y/o datos de referencia (por ejemplo, tablas que muestran valores de salida por defecto para el circuito electrónico dado, etc.). Por ejemplo, a partir de todos los criterios de búsqueda de la pluralidad de consultas de búsqueda y los datos de referencia, puede seleccionarse el conjunto de características que va a usarse por el algoritmo de aprendizaje de máquina. El conjunto de características puede contener por lo tanto una mezcla de criterios de búsqueda y características que requieren cálculos, pero una característica puede corresponder también a uno o más de búsqueda (por ejemplo, todos los parámetros de entrada para un circuito electrónico dado que va a simularse mediante las simulaciones informáticas de circuitos electrónicos, indicación de tiempo de la consulta de búsqueda, etc.). El conjunto de características están estructuradas ya que el algoritmo de aprendizaje de máquina requiere la entrada de, por ejemplo, un fichero de datos estructurado (por ejemplo, fichero de datos separado por comas), con una fila por consulta de búsqueda y una columna por característica. El orden de las columnas sigue un patrón dado para que se use el modelo probabilístico correctamente una vez desplegado. El algoritmo de aprendizaje de máquina puede crear también iterativamente un árbol de decisión. Empezando desde la pluralidad de las consultas de búsqueda seleccionadas, el algoritmo de aprendizaje de máquina prueba características y condiciones en estas características para dividir la pluralidad en dos conjuntos hijos. Una de estas pruebas puede ser "es el parámetro de entrada para el circuito electrónico dado igual a una tensión que tiene un valor de 10" (es decir, en general, comprobar si la estructura equivale a un valor dado), o "es el parámetro de entrada para el circuito electrónico dado igual a una tensión que tiene un valor mayor que 20" (es decir, en general, comprobar si la característica es mayor/menor que el valor dado). A partir de todas las características

y condiciones probadas, el algoritmo de aprendizaje de máquina mantiene únicamente la característica y condición que separa la mayoría de la pluralidad en un primer conjunto hijo que contiene consultas de búsqueda con su objetivo igual a un valor uno, y un segundo conjunto hijo que contiene las consultas de búsqueda con su objetivo no igual al valor uno. En otras palabras, el árbol de decisión se crea mediante el algoritmo de aprendizaje de máquina de manera que se toma una característica como un nodo con dos rutas de salida, una para consultas de búsqueda que tienen la característica que equivale al valor dado (el primer conjunto hijo) y otra ruta para consultas de búsqueda que tienen la característica que no es igual al valor dado (el segundo conjunto hijo). Es decir, las consultas de búsqueda se dividen en dos conjuntos hijos basándose en la característica. El algoritmo de aprendizaje de máquina crea progresivamente el árbol de decisión, dividiendo a continuación también los conjuntos hijos en conjuntos más pequeños con la misma característica y lógica de selección de condición. El algoritmo de aprendizaje de máquina se detiene una vez que el árbol de decisión alcanza un tamaño o complejidad dados proporcionados en los parámetros de los mismos.

Como se ha descrito anteriormente, la Figura 5 muestra operaciones que se realizan para cada consulta de búsqueda seleccionada en la operación 520 de entre la pluralidad de consultas de búsqueda manejadas por el sistema de base de datos 600. En el presente documento a continuación, las operaciones que van a realizarse para cada consulta de búsqueda que no se seleccionan en la operación 520 de la Figura 5 se describirán con referencia al método 700 mostrado en la Figura 7.

En respuesta a que no se haya seleccionado la consulta de búsqueda en la operación 520 (es decir, para la consulta de búsqueda que no se ha seleccionado aleatoriamente), se recupera una pieza de datos almacenados en caché (denominada como segunda pieza de datos almacenados en caché) que coincide con la consulta de búsqueda desde la fuente de caché 115 en la operación 710, similar a las operaciones 220, 225, y 230 en la Figura 2. Específicamente, se proporciona a la fuente de caché 115 con la solicitud de búsqueda como en la operación 220 de la Figura 2 y se buscan datos almacenados en caché que coinciden con la consulta de búsqueda, particularmente los criterios o parámetros de búsqueda de la misma, como en la operación 225 de la Figura 2. Se devuelven los datos almacenados en caché que coinciden con la consulta de búsqueda, como la segunda pieza de datos almacenados en caché, desde la fuente de caché 115 como en la operación 230 de la Figura 2.

La segunda pieza de datos almacenados en caché recuperados desde la fuente de caché 115 está asociada con el valor de precisión, como se ha descrito anteriormente. En respuesta a la recuperación de la segunda pieza de datos almacenados en caché en la operación 710, se evalúa el valor de precisión de la segunda pieza de datos almacenados en caché en la operación 720 por el pronosticador 650 como se muestra en la Figura 6. Por ejemplo, el valor de precisión se compara con un umbral dado. Esta operación es similar a la operación 240 mostrada en la Figura 2.

En algunos ejemplos, el umbral está prescrito por el cliente 130. Por ejemplo, el cliente 130 incluye el umbral (tal como "al menos el 85 %" o "al menos 0,9" o "alto" que se define análogamente como "al menos 0,9") en la consulta de búsqueda cuando se solicitan resultados de búsqueda del sistema de base de datos 600. El cliente 130 puede enviar también mensajes asíncronos especializados que indican un umbral deseado al sistema de base de datos 600. El sistema de base de datos 600 almacena estas prescripciones de umbral específicas de cliente y las emplea cada vez que se recibe una consulta de búsqueda desde el respectivo cliente 130. Como alternativa, en otros ejemplos, el umbral se establece por una tercera parte tal como el operador del sistema de base de datos 600. En este caso, un único umbral puede ser aplicable para todas las consultas de búsqueda recibidas de todos los clientes 130. Los clientes 130 pueden no tener una influencia sobre el umbral empleado por el sistema de base de datos 600. Como alternativa, el umbral preestablecido por la tercera parte puede actuar como un valor por defecto y los clientes 130 pueden anular este valor por defecto mediante una prescripción específica de cliente propia.

Si la evaluación en la operación 720 resulta que el valor de precisión se encuentra por debajo del umbral dado, se recupera una pieza de datos originales (denominada como segunda pieza de datos originales) que coincide con la consulta de búsqueda desde la fuente de datos original 120 en la operación 730. Esta operación 730 es similar a las operaciones 250, 255, y 260 mostradas en la Figura 2. Más específicamente, se proporciona la fuente de datos original 120 con la solicitud de búsqueda, similar a la operación 250 en la Figura 2, y se buscan datos originales que coinciden con la consulta de búsqueda, particularmente los criterios o parámetros de búsqueda de la misma, de manera similar a la operación 255 en la Figura 2. Se devuelven los datos originales que coinciden con la consulta de búsqueda, como la segunda pieza de datos originales, desde la fuente de datos original 120, de manera similar a la operación 260 en la Figura 2.

En respuesta a la recuperación de la segunda pieza de datos originales, se actualiza la segunda pieza de datos almacenados en caché en la fuente de caché 115 por la segunda pieza de datos originales en la operación 740. Esta operación 740 es similar a la operación 270 mostrada en la Figura 2. Por ejemplo, la segunda pieza de datos originales se almacena en la fuente de caché 115 de manera que se sustituye la segunda pieza de datos almacenados en caché, formando de esta manera una nueva pieza de datos almacenados en caché en la fuente de caché 115. También, la segunda pieza de datos originales puede almacenarse en la fuente de caché 115 como la pieza actualizada de datos almacenados en caché, que invalida la segunda pieza de datos almacenados en caché anterior. Pueden actualizarse también los datos de control relacionados con el valor de precisión de la segunda pieza de datos almacenados en caché actualizada tal como una indicación de tiempo de actualización. En un ejemplo, puede realizarse un colector de basura en la fuente de caché 115, que borra cualesquiera datos almacenados en caché invalidados y que libera el

correspondiente espacio en la fuente de caché 115.

Dependiendo de si la evaluación en la operación 720 resulta que su valor de precisión se encuentra o no por encima del umbral dado, se devuelve la segunda pieza de datos almacenados en caché o la segunda pieza de datos originales como el resultado de búsqueda al cliente 130 en la operación 750. Esta operación 750 puede realizarse también en paralelo con la operación 740. En este punto, en caso de que se hayan hecho menos restrictivos los criterios o parámetros de búsqueda de la consulta de búsqueda como se ha descrito anteriormente, devolver la segunda pieza de datos originales o la segunda pieza de datos almacenados en caché incluye una indicación de cuáles y cómo se han cambiado los criterios o parámetros de búsqueda de la consulta de búsqueda de modo que son menos restrictivos.

En un ejemplo, la operación 550 del método 500 mostrado en la Figura 5 puede realizarse de acuerdo con el método 800 mostrado en la Figura 8.

En respuesta a la recuperación de la primera pieza de datos almacenados en caché desde la fuente de caché 115 y la recuperación de la primera pieza de datos originales desde la fuente de datos original 120, puede determinarse si actualizar la primera pieza de datos almacenados en caché es o no apropiado.

El método 800 incluye, en la operación 810, comparar la primera pieza de datos almacenados en caché y la primera pieza de datos originales. La operación 810 de comparación se realiza para comprobar si la primera pieza de datos almacenados en caché recuperados desde la fuente de caché 115 coincide con la primera pieza de datos originales recuperados desde la fuente de datos original 120. Si se determina que las primeras piezas de datos almacenados en caché/originales no coinciden, que significa que la primera pieza de datos almacenados en caché ya se ha desactualizado, se actualiza la primera pieza de datos almacenados en caché por la primera pieza de datos originales, en la operación 550. Esta operación 550 es similar a la operación 270 mostrada en la Figura 2. De manera independiente del resultado de la operación 810, la primera pieza de datos originales se devuelve al cliente 130, desde el cual se ha recibido la consulta de búsqueda, como el resultado de búsqueda.

En un ejemplo, en el selector 620 o en el sistema de base de datos 600, puede establecerse un número de consultas de búsqueda que va a seleccionarse de entre todas las consultas de búsqueda manejadas por el sistema de base de datos 600. Esto aumenta la importancia probabilística ya que se considera el número de consultas de búsqueda seleccionadas al adaptar el modelo de probabilidad. Este número de consultas de búsqueda seleccionadas puede ser específico de sistema, estar predefinido, o determinado basándose en un factor derivado del modelo probabilístico. El número de consultas de búsqueda seleccionadas puede aplicarse para un periodo de tiempo dado tal como, por ejemplo, una hora o un día. Por consiguiente, la operación 520 mostrada en la Figura 5 puede adaptarse para posibilitar la selección de consultas de búsqueda de entre las consultas de búsqueda manejadas por el sistema de base de datos 600 en un número que corresponde al número de consultas de búsqueda establecidas seleccionadas. El selector 620 realiza la operación 520 únicamente en caso de que un número de consultas de búsqueda previamente seleccionadas no sea igual al número de consultas de búsqueda seleccionadas establecido; de lo contrario las consultas de búsqueda se manejan de acuerdo con el método 700 mostrado en la Figura 7. En caso de selección de una consulta de búsqueda, el selector 420 incrementa el número de consultas de búsqueda previamente seleccionadas. El selector 420 puede resetear el número de consultas de búsqueda previamente seleccionadas después de un periodo de tiempo dado para el que se aplica el número de consultas de búsqueda seleccionadas.

En otro ejemplo, el sistema de base de datos 600 incluye una base de datos de historial o de registro 630. La base de datos de historial o de registro 630 puede realizarse mediante cualquier clase de bases de datos, preferentemente mediante una base de datos distribuida, y puede estar separada de la fuente de datos original 120 o combinada con la misma. En la base de datos de historial o de registro 630, se almacena información acerca de consultas de búsqueda, datos almacenados en caché y originales recuperados en respuesta a consultas de búsqueda y se mantiene para procesamiento adicional. En el método 500 mostrado en la Figura 5, por ejemplo, después de la operación 540 para recuperar la primera pieza de datos originales desde la fuente de datos original 120, se envía información acerca de la consulta de búsqueda recibida desde el cliente 130, la primera pieza de datos almacenados en caché recuperada desde la fuente de caché 115 en la operación 530, y la primera pieza de datos originales recuperada desde la fuente de datos original 120 en la operación 540 a la base de datos de historial o de registro 630. Por ejemplo, la información acerca de la consulta de búsqueda puede incluir uno o más de lo siguiente: los criterios o parámetros de búsqueda de la consulta de búsqueda, el tiempo de recepción de la consulta de búsqueda en el sistema de base de datos 600, información acerca del cliente 130 desde el que se ha recibido la consulta de búsqueda, e información acerca de los tiempos de respuesta para recuperar las primeras piezas de datos almacenados en caché/originales. También, la primera pieza de datos almacenados en caché puede enviarse a la base de datos de historial o de registro 630 junto con el valor de precisión con el que está asociada. La base de datos de historial o de registro 630 recibe la información y almacena la misma en una manera estructurada, por ejemplo, como una n-tupla (por ejemplo, 3-tupla o triple). La base de datos de historial o de registro 630 mantiene y proporciona la información al modelador 610, que usa la misma al adaptar el modelo probabilístico en la operación 560. La información puede mantenerse en la base de datos de historial o de registro 630 durante un periodo de tiempo dado, o descartarse una vez usada al adaptar el modelo probabilístico en la operación 560.

En otro ejemplo más, el sistema de base de datos 600 incluye un actualizador de umbral 640. El actualizador de umbral

640 es responsable de actualizar el umbral dado, si se emplea mediante el sistema de base de datos 600. La adaptación o variación del umbral dado posibilita que el sistema de base de datos 600 controle la carga de manera que únicamente se procese un número o parte de las consultas de búsqueda manejadas por el sistema de base de datos 600 usando la fuente de datos original 120, evitando de esta manera la sobrecarga en la fuente de datos original 120. También, el actualizador de umbral 640 puede proporcionar y actualizar una regla para adaptar el umbral prescrito por los clientes 130. El actualizador de umbral 640 realiza operaciones preferentemente de una manera asíncrona con las operaciones mostradas en la Figura 5, así como las operaciones del modelador 610 y la base de datos de historial 630. El actualizador de umbral 640 actualiza el umbral dado en una base regular, si se activa por un tercero, o si el número de consultas de búsqueda manejadas como consultas de búsqueda seleccionadas supera un valor dado.

La operación del actualizador de umbral 640 se describirá ahora con referencia a la Figura 9 que muestra un método 900 de actualización del umbral dado. Preferentemente, el método 900 de actualización del umbral dado es parte del método 500 de adaptación del modelo probabilístico como se muestra en la Figura 5.

El método 900 de la Figura 9 se inicia con la comparación de la primera pieza de datos originales con la primera pieza de datos almacenados en caché en la operación 910. La primera pieza de datos originales con la primera pieza de datos almacenados en caché que va a compararse puede recuperarse por el actualizador de umbral 640 desde la base de datos de historial o de registro 630. También, las primeras piezas de datos almacenados en caché/originales que van a compararse pueden enviarse al actualizador de umbral 640 en respuesta a la recuperación de las mismas desde la fuente de caché 115 y la fuente de datos original 120.

En la operación 920, se almacena un primer valor indicador por el actualizador de umbral 640. El primer valor indicador se proporciona por el actualizador de umbral 640 para las primeras piezas de datos almacenados en caché/originales comparadas, e indica el resultado de la comparación. Es decir, el primer valor indicador es un indicador de si las primeras piezas de datos almacenados en caché/originales coinciden o no entre sí. Por ejemplo, el primer valor indicador se representa por un valor de un bit, donde el valor "1" representa que las primeras piezas coinciden entre sí, mientras que el valor "0" representa que las primeras piezas no coinciden. El primer valor indicador puede mantenerse en el actualizador de umbral 640 en asociación con las primeras piezas de datos almacenados en caché/originales o puede enviarse a la base de datos de historial o de registro 630 para su almacenamiento como un valor en la n-tupla de las correspondientes primeras piezas de datos almacenados en caché/originales.

A continuación, en la operación 930, el valor de precisión asociado con la primera pieza de datos almacenados en caché recuperado de la base de datos de historial o de registro 630 se compara con el umbral dado, es decir, con el umbral dado actualmente establecido en el sistema de base de datos 600 o en vista de la correspondiente consulta de búsqueda.

En la operación 940, se almacena un segundo valor indicador por el actualizador de umbral 640. El segundo valor indicador se proporciona por el actualizador de umbral 640 para la primera pieza de datos almacenados en caché, el valor de precisión de la cual se ha usado en la comparación de la operación 930, e indica el resultado de la comparación. Es decir, el segundo valor indicador es un valor indicador de si el valor de precisión se encuentra o no por debajo del umbral dado. Por ejemplo, el segundo valor indicador se representa por un valor de un bit, donde el valor "1" representa que el valor de precisión se encuentra por debajo del umbral dado, mientras que el valor "0" representa que el valor de precisión no se encuentra por debajo del umbral dado. El segundo valor indicador puede mantenerse en el actualizador de umbral 640 en asociación con la primera pieza de datos almacenados en caché o puede enviarse a la base de datos de historial o de registro 630 para su almacenamiento como valor en la n-tupla de la correspondiente primera pieza de datos almacenados en caché.

Basándose en el primer y segundo valores indicadores, mantenidos en el actualizador de umbral 640 o recuperados desde la base de datos de historial o de registro 630, el umbral dado se actualiza en la operación 940.

El umbral dado según se actualiza por el método 900 se establece en el sistema de base de datos 600 y se usa para posteriores consultas de búsqueda.

En un ejemplo preferido, se realiza el método 900 de la Figura 9 por el actualizador de umbral 640 para cada una de una pluralidad de primeras piezas de datos almacenados en caché/originales almacenados en la base de datos de historial o de registro 630 en forma de procesamiento por lotes.

En un ejemplo, la operación 940 del método 900 incluye la operación 942 de determinación de una tasa de interrogación y una eficacia de interrogación basándose en pares del primer y segundo valores indicadores, y la operación 944 de actualización del umbral dado basándose en la tasa de interrogación y la eficacia de interrogación. En este punto, se tiene en cuenta una pluralidad del primer y segundo valores indicadores, y se recuperan posiblemente de la base de datos de historial o de registro 630 antes de la operación 942. La tasa de interrogación indica la tasa a la que se realiza la operación de interrogación. Es decir, la tasa de interrogación es la tasa entre el número total de operaciones de interrogación realmente realizadas y el número total de posibles operaciones de interrogación (es decir, si se realiza la operación de interrogación para cada consulta de búsqueda). Por ejemplo, la tasa de interrogación es una relación entre un número de los segundos valores indicadores que indican que el valor

de precisión se encuentra por debajo del umbral dado, es decir, que indica que la primera pieza de datos almacenados en caché se ha desactualizado y se asume que no coincide con la correspondiente primera pieza de datos originales, y un número total de los segundos valores indicadores. La eficacia de interrogación indica cómo de eficaz han sido las operaciones de interrogación. Es decir, la eficacia de interrogación es una relación entre el número de operaciones de interrogación realizadas mientras el valor de precisión se encuentra por debajo el umbral dado y los datos almacenados en caché que ya se han desactualizado, y el número total de operaciones de interrogación realmente realizadas. Por ejemplo, la eficacia de interrogación es una relación entre un número de los segundos valores indicadores que indican que el valor de precisión se encuentra por debajo del umbral dado, en donde la primera pieza de datos almacenados en caché ya se ha desactualizado, y un número total de los segundos valores indicadores que indican que el valor de precisión se encuentra por debajo del umbral dado.

En el ejemplo anterior, la tasa de interrogación puede determinarse basándose en la siguiente ecuación (1):

$$\text{Tasa de interrogación} = \frac{(TP+FP)}{(TP+FP+TN+FN)} \quad (1)$$

la eficacia de interrogación puede determinarse en el ejemplo anterior basándose en la siguiente ecuación (2):

$$\text{Eficacia de interrogación} = \frac{TP}{(TP+FP)} \quad (2)$$

En las ecuaciones (1) y (2):

- TP es un número de dichos pares en donde el primer valor indicador indica que la pieza de datos originales no coincide con la pieza de datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión se encuentra por debajo del umbral dado;
- TN es un número de dichos pares en donde el primer valor indicador indica que la pieza de datos originales coincide con la pieza de datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión no se encuentra por debajo del umbral dado;
- FP es un número de dichos pares en donde el primer valor indicador indica que la pieza de datos originales coincide con la pieza de datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión se encuentra por debajo del umbral dado, y
- FN es un número de dichos pares en donde el primer valor indicador indica que la pieza de datos originales no coincide con la pieza de datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión no se encuentra por debajo del umbral dado.

También, en el ejemplo anterior, si la tasa de interrogación calculada en la pluralidad de pares se encuentra por encima de un valor deseado (por ejemplo, el límite de hardware de la fuente de datos original 120), puede aumentarse el umbral dado. de lo contrario, si la tasa de interrogación es inferior al valor deseado, puede reducirse el umbral dado. De manera iterativa en un lote o varios lotes de la pluralidad de pares, el umbral dado converge hacia un valor que le posibilita alcanzar el valor deseado de la tasa de interrogación.

Finalmente, la Figura 10 es una representación esquemática de un sistema informático que proporciona la funcionalidad del sistema de base de datos 600 de la Figura 6. Dentro del sistema de base de datos 600a puede ejecutarse un conjunto de instrucciones, para hacer que el sistema informático realice cualquiera de las metodologías analizadas en el presente documento. El sistema de base de datos 600 incluye al menos un procesador 1020, una memoria principal 1040 y un dispositivo de interfaz de red 1080, que se comunican entre sí mediante un bus 1010. Opcionalmente, puede incluir adicionalmente una memoria estática 1060 y una unidad de almacenamiento de disco 1070. Puede proporcionarse una pantalla de vídeo, un dispositivo de entrada alfanumérico y un dispositivo control de cursor como ejemplos de interfaz de usuario 1030. El dispositivo de interfaz de red 1080 conecta el sistema informático 600 a al menos los clientes 130, una o más bases de datos que actúan como la fuente de datos original 120 y la base de datos de historial o de registro 630, si se proporciona externa al sistema de base de datos 600, otros dispositivos informáticos, si se proporcionan externos al sistema de base de datos 600, que actúan como el modelador 610, el selector 620 y el actualizador de umbral 640, Internet y/o cualquier otra red. También, la una o más bases de datos que actúan como la fuente de datos original 120 y la base de datos de historial o de registro 630 pueden proporcionarse como bases de datos 1042 dentro de la memoria principal 1040. Un conjunto de instrucciones ejecutables por ordenador (es decir, código de programa informático) 1044 que realizan una, o todas, las metodologías anteriormente descritas, residen completamente, o al menos parcialmente, dentro de o en un medio legible por máquina, por ejemplo, la memoria principal 1040 y/o el al menos un procesador 1020. Un medio legible por máquina en el que reside el código 1044 puede ser también un soporte de datos no volátil (por ejemplo un disco duro magnético no extraíble o un disco óptico o magnético extraíble) que es parte de la unidad de control de disco 1070. El código 1044 puede transmitirse adicionalmente o recibirse como una señal propagada mediante la Internet a través del dispositivo de interfaz de red 1080. La operación básica del sistema de base de datos 600 que incluye la interfaz de usuario y la comunicación de red se controla por el sistema operativo 1046.

En general, las rutinas ejecutadas para implementar las realizaciones de la invención, ya se implementen como parte de un sistema operativo o una aplicación, componente, programa, objeto, módulo o secuencia de instrucciones

específico, o incluso un subconjunto de los mismos, se pueden denominar en el presente documento "código de programa informático" o simplemente "código de programa". Código de programa habitualmente comprende instrucciones legibles por ordenador que están residentes en diversas veces en diversos dispositivos de memoria y de almacenamiento en un ordenador y que, cuando se leen y ejecutan por uno o más procesadores en un ordenador, provocan que ese ordenador realice las operaciones necesarias para ejecutar operaciones y/o elementos que incorporan los diversos aspectos de las realizaciones de la invención. Instrucciones de programa legibles por ordenador para efectuar operaciones de las realizaciones de la invención pueden ser, por ejemplo, lenguaje de ensamblaje o bien código fuente o bien código objeto escrito en cualquier combinación de uno o más lenguajes de programación.

Diverso código de programa descrito en este documento puede identificarse basándose en la aplicación dentro de la que se implementa en realizaciones específicas de la invención. Sin embargo, debería apreciarse que cualquier nomenclatura de programa particular a continuación se usa meramente para conveniencia y, por lo tanto, la invención no debería limitarse a usar solamente en cualquier aplicación especificada identificada y/o implícita por tal nomenclatura. Adicionalmente, dado el número generalmente infinito de maneras en las que pueden organizarse programas informáticos en rutinas, procedimientos, métodos, módulos, objetos y similares, así como las diversas maneras en las que la funcionalidad de programa puede asignarse entre diversas capas de software que están residentes dentro de un ordenador típico (por ejemplo, sistemas operativos, librerías, API, aplicaciones, subprogramas, etc.), debería apreciarse que las realizaciones de la invención no se limitan a la organización específica y asignación de funcionalidad de programa descritas en este documento.

El código de programa incorporado en cualquiera de las aplicaciones/módulos descritos en este documento se puede distribuir individual o colectivamente como un producto de programa en una diversidad de diferentes formas. En particular, el código de programa puede distribuirse usando un medio de almacenamiento legible por ordenador que tienen instrucciones de programa legibles por ordenador en el mismo para provocar que un procesador efectúe aspectos de las realizaciones de la invención.

Medio de almacenamiento legible por ordenador, que es inherentemente no transitorio, puede incluir medio tangible volátil y no volátil y extraíble y no extraíble implementado en cualquier método o tecnología para almacenamiento de información, tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programa, u otros datos. Medio de almacenamiento legible por ordenador puede incluir adicionalmente memoria de acceso aleatorio (RAM), memoria de sólo lectura (ROM), memoria de sólo lectura programable borrable (EPROM), memoria de sólo lectura eléctricamente programable borrable (EEPROM), memoria flash u otra tecnología de memoria de estado sólido, memoria de sólo lectura de disco compacto portátil (CD-ROM), u otro almacenamiento óptico, cintas magnéticas, cinta magnética, almacenamiento de disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio que pueda usarse para almacenar la información deseada y que pueda leerse por un ordenador. Un medio de almacenamiento legible por ordenador no debería interpretarse como señales transitorias en sí (por ejemplo, ondas de radio u otras ondas electromagnéticas que se propagan, ondas electromagnéticas que se propagan a través de un medio de transmisión tal como una guía de ondas, o señales eléctricas transmitidas a través de un alambre). Instrucciones de programa legibles por ordenador pueden descargarse a un ordenador, otro tipo de aparato de procesamiento de datos programable, u otro dispositivo desde un medio de almacenamiento legible por ordenador o a un ordenador externo o dispositivo de almacenamiento externo a través de una red.

Pueden usarse instrucciones de programa legibles por ordenador almacenadas en un medio legible por ordenador para dirigir un ordenador, otros tipos de aparato de procesamiento de datos programable, u otros dispositivos para funcionar de una manera particular, de manera que las instrucciones almacenadas en el medio legible por ordenador producen un artículo de fabricación que incluye instrucciones que implementan las funciones, actos y/u operaciones especificadas en los diagramas de flujo, diagramas de secuencia y/o diagramas de bloques. Las instrucciones de programa informáticas pueden proporcionarse a uno o más procesadores de un ordenador de fin general, un ordenador de fin especial, u otro aparato de procesamiento de datos programable para producir una máquina, de tal forma que las instrucciones, que se ejecutan a través del uno o más procesadores, provocan que se realicen una serie de cálculos para implementar las funciones, actos y/u operaciones especificados en el diagrama de flujos, diagrama de secuencias y/o diagramas de bloque.

En ciertas realizaciones alternativas, las funciones, actos y/u operaciones especificados en el diagrama de flujos, diagrama de secuencias y/o diagramas de bloque pueden reordenarse, procesarse en serie y/o procesarse simultáneamente consistentes con realizaciones de la invención. Además, cualquiera del diagrama de flujos, diagrama de secuencias y/o diagramas de bloque puede incluir más o menos bloques que los ilustrados consistentes con realizaciones de la invención.

La terminología usada en este documento es para el propósito de describir únicamente realizaciones particulares y no se pretende que sea limitante de las realizaciones de la invención. Se entenderá adicionalmente que los términos "comprende" y/o "que comprende", cuando se usan en esta memoria descriptiva, especifican la presencia de características indicadas, elementos integrantes, etapas, operaciones, elementos, y/o componentes, pero no excluyen la presencia o adición de una o más otras características, elementos integrantes, etapas, operaciones, elementos, componentes, y/o grupos de los mismos. Asimismo, en la medida en la que las expresiones "incluye", "teniendo/que

tiene", "tiene", "con", "compuesto por" o variantes de las mismas se usan o bien en la descripción detallada o bien en las reivindicaciones, se pretende que tales expresiones sean inclusivas de una manera similar a la expresión "comprendiendo/que comprende".

- 5 Aunque la totalidad de la invención se ha ilustrado mediante una descripción de diversas realizaciones, y aunque estas realizaciones se han descrito con un detalle considerable, la intención no es restringir o limitar en modo alguno el alcance de las reivindicaciones adjuntas a tal detalle.

REIVINDICACIONES

1. Un método de procesamiento de consultas en un sistema de base de datos, comprendiendo el sistema de base de datos una fuente de datos original que almacena una pluralidad de piezas de datos originales y una fuente de caché que almacena una pluralidad de piezas de datos almacenados en caché, estando asociada cada pieza de datos almacenados en caché con un valor de precisión, comprendiendo el método:
5 seleccionar aleatoriamente consultas de entre una pluralidad de consultas manejadas por el sistema de base de datos, en el momento que se recibe la respectiva consulta desde un cliente;
10 para cada una de las consultas seleccionadas aleatoriamente,

 recuperar una primera pieza de los datos almacenados en caché que coincide con la consulta seleccionada aleatoriamente desde la fuente de caché, y
15 recuperar una primera pieza de los datos originales que coincide con la consulta seleccionada aleatoriamente desde la fuente de datos original;

 para consultas que no se seleccionan aleatoriamente,

20 recuperar una segunda pieza de los datos almacenados en caché que coincide con la consulta desde la fuente de caché;
 evaluar el valor de precisión de la segunda pieza de los datos almacenados en caché;
 si el valor de precisión se encuentra por debajo de un umbral dado,

25 recuperar una segunda pieza de los datos originales que coincide con la consulta desde la fuente de datos original, y
 actualizar la segunda pieza de los datos almacenados en caché mediante la segunda pieza de los datos originales; y

30 adaptar un modelo probabilístico basándose en la primera pieza de los datos almacenados en caché y la primera pieza de los datos originales usando un algoritmo de aprendizaje de máquina,
 en donde se deriva el valor de precisión del modelo probabilístico e indica una probabilidad de que una pieza de los datos almacenados en caché coincida con una correspondiente pieza de los datos originales.
- 35 2. El método de la reivindicación 1, que comprende adicionalmente: para las consultas seleccionadas aleatoriamente,

 comparar la primera pieza de los datos almacenados en caché y la primera pieza de los datos originales;
 actualizar la primera pieza de los datos almacenados en caché mediante la primera pieza de los datos originales, y devolver la primera pieza de los datos originales al cliente.
- 40 3. El método de la reivindicación 1 o 2, que comprende adicionalmente: para consultas que no se seleccionan aleatoriamente,

 si el valor de precisión se encuentra por debajo del umbral dado, devolver la segunda pieza de los datos originales como resultado al cliente; y
45 de lo contrario, devolver la segunda pieza de los datos almacenados en caché como resultado al cliente.
- 50 4. El método de una cualquiera de las reivindicaciones 1 a 3, que comprende adicionalmente:
 para cada una de las consultas seleccionadas aleatoriamente,
 enviar información acerca de la consulta, la primera pieza de los datos almacenados en caché y la primera pieza de los datos originales a una base de datos de historial del sistema de base de datos.
- 55 5. El método de una cualquiera de las reivindicaciones 1 a 4, que comprende adicionalmente:

 comparar la primera pieza de los datos originales con la primera pieza de los datos almacenados en caché;
 almacenar un primer valor indicador que indica si la primera pieza de los datos originales coincide con la primera pieza de los datos almacenados en caché;
 comparar el valor de precisión de la primera pieza de los datos almacenados en caché y el umbral dado;
 almacenar un segundo valor indicador que indica si el valor de precisión se encuentra por debajo del umbral dado;
60 y
 actualizar el umbral dado basándose en pares del primer y el segundo valores indicadores.
- 65 6. El método de la reivindicación 5, en donde actualizar el umbral dado comprende:

 determinar una tasa de interrogación y una eficacia de interrogación basándose en los pares del primer y segundo valores indicadores; y
 actualizar el umbral dado basándose en la tasa de interrogación y la eficacia de interrogación.

7. El método de la reivindicación 6, en donde la tasa de interrogación se determina basándose en la ecuación: $(TP+FP) / (TP+FP+TN+FN)$, y la eficacia de interrogación se determina basándose en la ecuación: $TP / (TP+FP)$, donde:

TP es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales no coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión se encuentra por debajo del umbral dado,

TN es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión no se encuentra por debajo del umbral dado,

FP es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión se encuentra por debajo del umbral dado, y

FN es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales no coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de precisión no se encuentra por debajo del umbral dado.

8. Un sistema de base de datos de procesamiento de consultas, que comprende una fuente de datos original que almacena una pluralidad de piezas de datos originales y una fuente de caché que almacena una pluralidad de piezas de datos almacenados en caché, estando asociada cada pieza de datos almacenados en caché con un valor de precisión, el sistema de base de datos configurado para:

seleccionar aleatoriamente consultas de entre una pluralidad de consultas manejadas por el sistema de base de datos, en el momento que se recibe la respectiva consulta desde un cliente; para cada una de las consultas seleccionadas aleatoriamente,

recuperar una primera pieza de los datos almacenados en caché que coincide con la consulta seleccionada aleatoriamente desde la fuente de caché, y recuperar una primera pieza de los datos originales que coincide con la consulta seleccionada aleatoriamente desde la fuente de datos original;

para consultas que no se seleccionan aleatoriamente,

recuperar una segunda pieza de los datos almacenados en caché que coincide con la consulta desde la fuente de caché; evaluar el valor de precisión de la segunda pieza de los datos almacenados en caché; si el valor de precisión se encuentra por debajo de un umbral dado,

recuperar una segunda pieza de los datos originales que coincide con la consulta desde la fuente de datos original, y actualizar la segunda pieza de los datos almacenados en caché mediante la segunda pieza de los datos originales; y

adaptar un modelo probabilístico basándose en la primera pieza de los datos almacenados en caché y la primera pieza de los datos originales usando un algoritmo de aprendizaje de máquina, en donde se deriva el valor de precisión del modelo probabilístico e indica una probabilidad de que una pieza de los datos almacenados en caché coincida con una correspondiente pieza de los datos originales.

9. El sistema de base de datos de la reivindicación 8, en donde el sistema de base de datos está configurado adicionalmente para:

para las consultas seleccionadas aleatoriamente,

comparar la primera pieza de los datos almacenados en caché y la primera pieza de los datos originales; actualizar la primera pieza de los datos almacenados en caché mediante la primera pieza de los datos originales, y devolver la primera pieza de los datos originales al cliente.

10. El sistema de base de datos de la reivindicación 8 o 9, en donde el sistema de base de datos está configurado adicionalmente para: para consultas que no se seleccionan aleatoriamente,

si el valor de precisión se encuentra por debajo del umbral dado, devolver la segunda pieza de los datos originales como resultado al cliente; y de lo contrario, devolver la segunda pieza de los datos almacenados en caché como resultado al cliente.

11. El sistema de base de datos de una cualquiera de las reivindicaciones 8 a 10, en donde el sistema de base de datos está configurado adicionalmente para:

para cada una de las consultas seleccionadas aleatoriamente,
 enviar información acerca de la consulta, la primera pieza de los datos almacenados en caché y la primera pieza
 de los datos originales a una base de datos de historial del sistema de base de datos.

5 12. El sistema de base de datos de una cualquiera de las reivindicaciones 8 a 11, en donde el sistema de base de
 datos está configurado adicionalmente para:

10 comparar la primera pieza de los datos originales con la primera pieza de los datos almacenados en caché;
 almacenar un primer valor indicador que indica si la primera pieza de los datos originales coincide con la primera
 pieza de los datos almacenados en caché;
 comparar el valor de precisión de la primera pieza de los datos almacenados en caché y el umbral dado;
 almacenar un segundo valor indicador que indica si el valor de precisión se encuentra por debajo del umbral dado;
 y
 15 actualizar el umbral dado basándose en pares del primer y el segundo valores indicadores.

13. El sistema de base de datos de la reivindicación 12, en donde el sistema de base de datos está configurado para
 actualizar el umbral dado:

20 determinando una tasa de interrogación y una eficacia de interrogación basándose en los pares del primer y
 segundo valores indicadores; y
 actualizando el umbral dado basándose en la tasa de interrogación y la eficacia de interrogación.

25 14. El sistema de base de datos de la reivindicación 13, en donde el sistema de base de datos está configurado para:
 la tasa de interrogación se determina basándose en la ecuación: $(TP+FP) / (TP+FP+TN+FN)$, y
 la eficacia de interrogación se determina basándose en la ecuación: $TP / (TP+FP)$,
 donde:

30 TP es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales
 no coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor
 de precisión se encuentra por debajo del umbral dado,
 TN es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales
 coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de
 35 precisión no se encuentra por debajo del umbral dado,
 FP es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales
 coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor de
 precisión se encuentra por debajo del umbral dado, y
 FN es un número de dichos pares en donde el primer valor indicador indica que la pieza de los datos originales
 40 no coincide con la pieza de los datos almacenados en caché, y el segundo valor indicador indica que el valor
 de precisión no se encuentra por debajo del umbral dado.

15. Un producto de programa informático que comprende instrucciones que, cuando se ejecutan por un ordenador,
 hacen que el ordenador realice el método de acuerdo con una cualquiera de las reivindicaciones 1 a 7.
 45

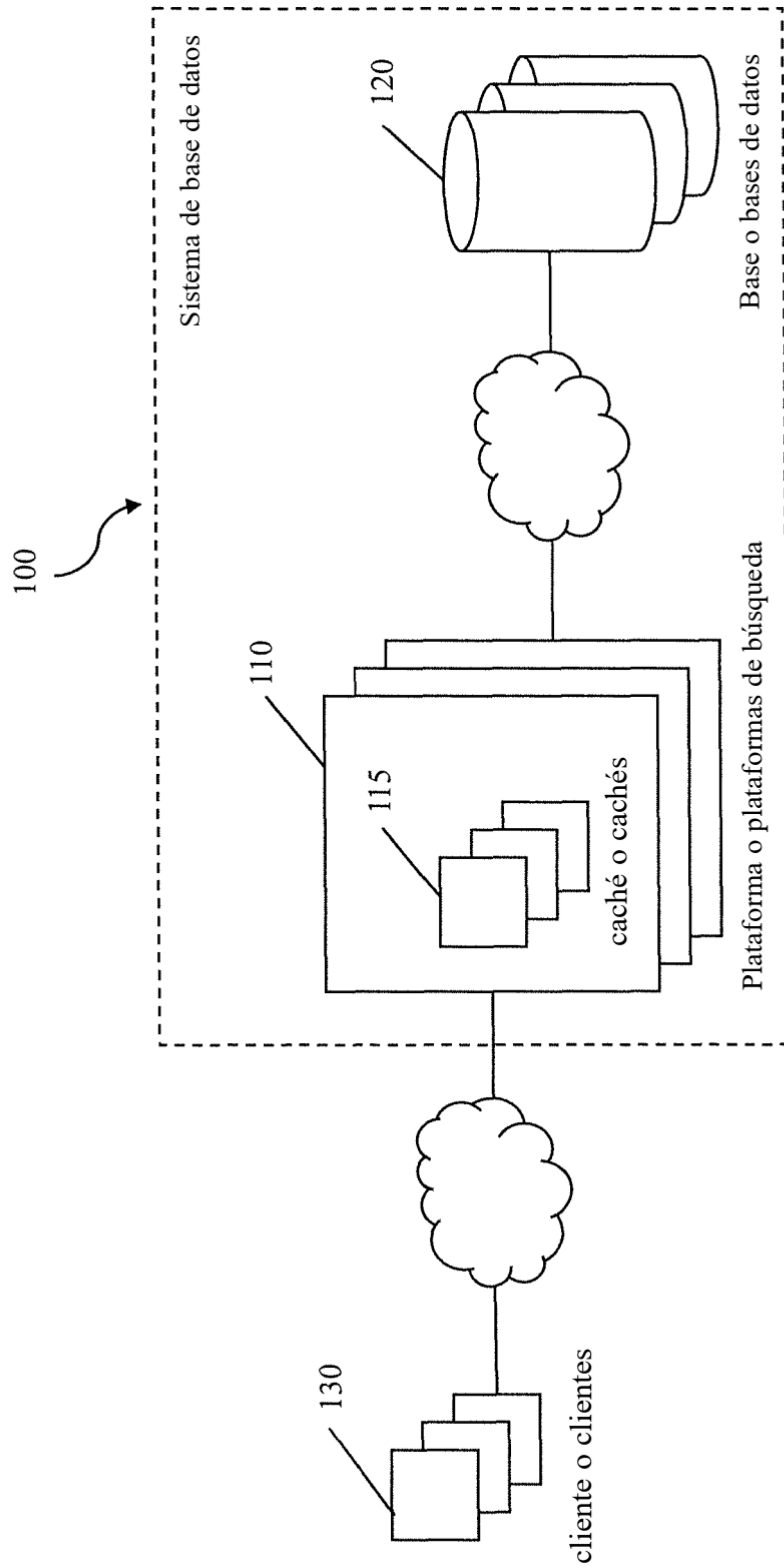


FIG. 1

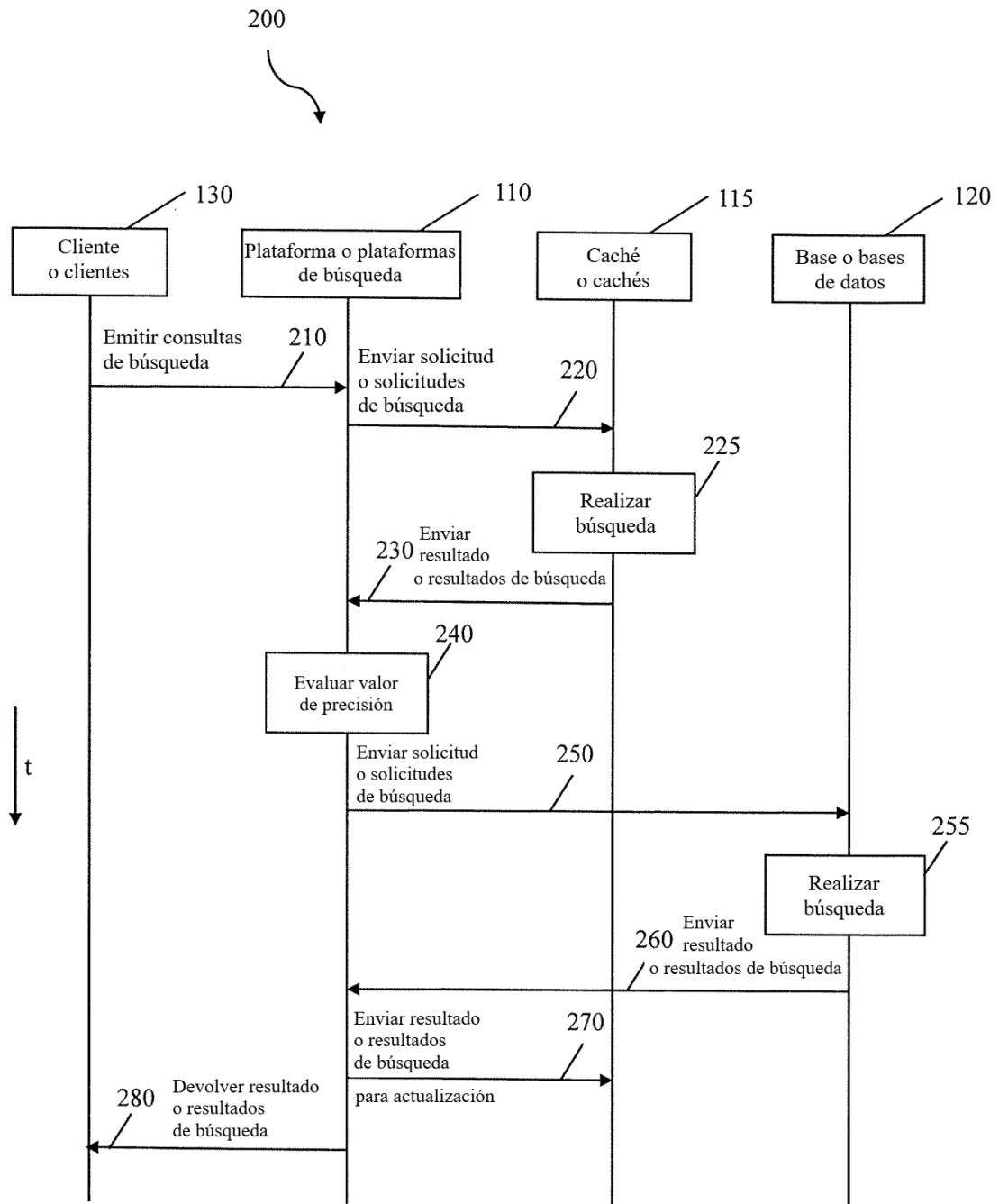


FIG. 2

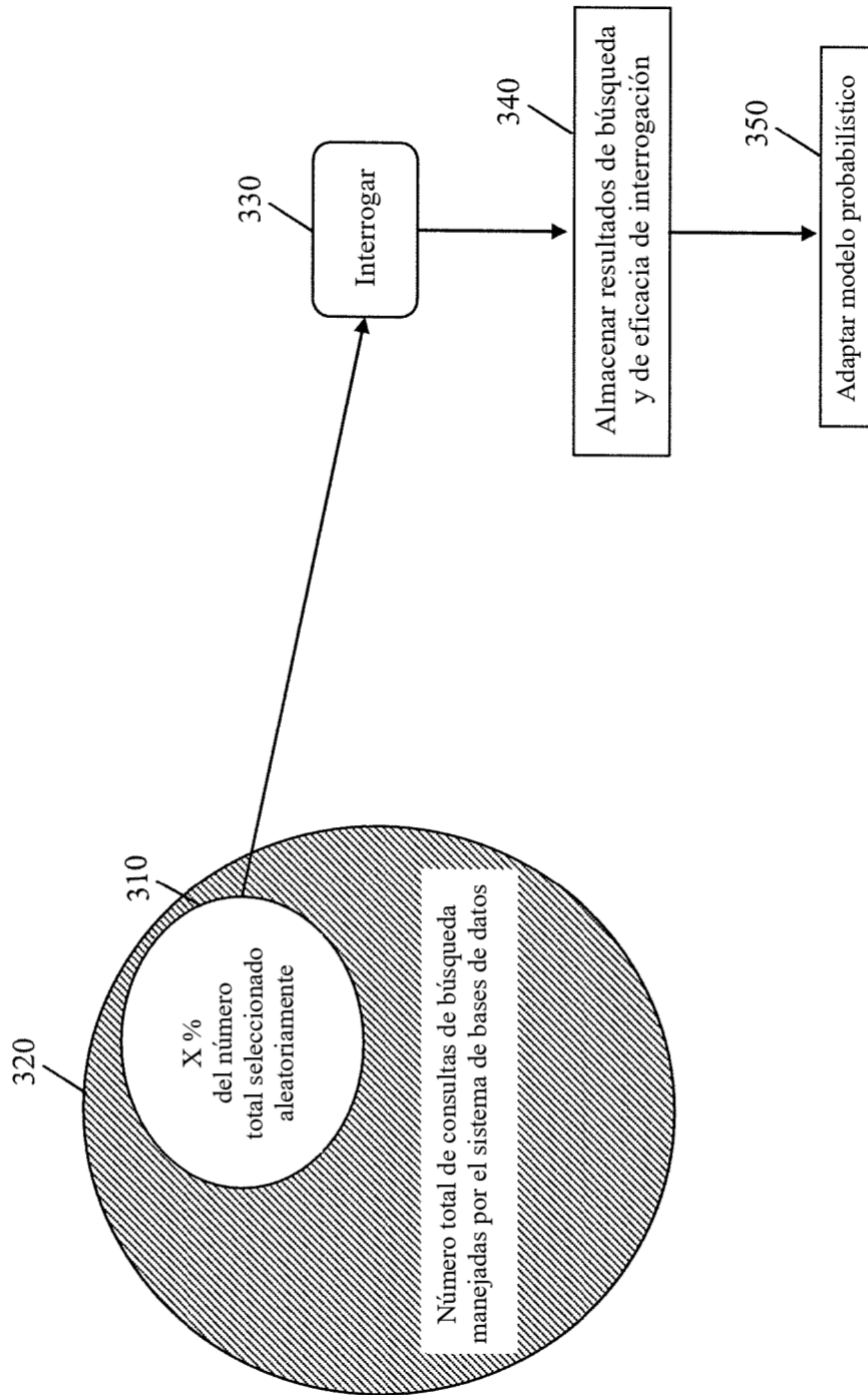


FIG. 3

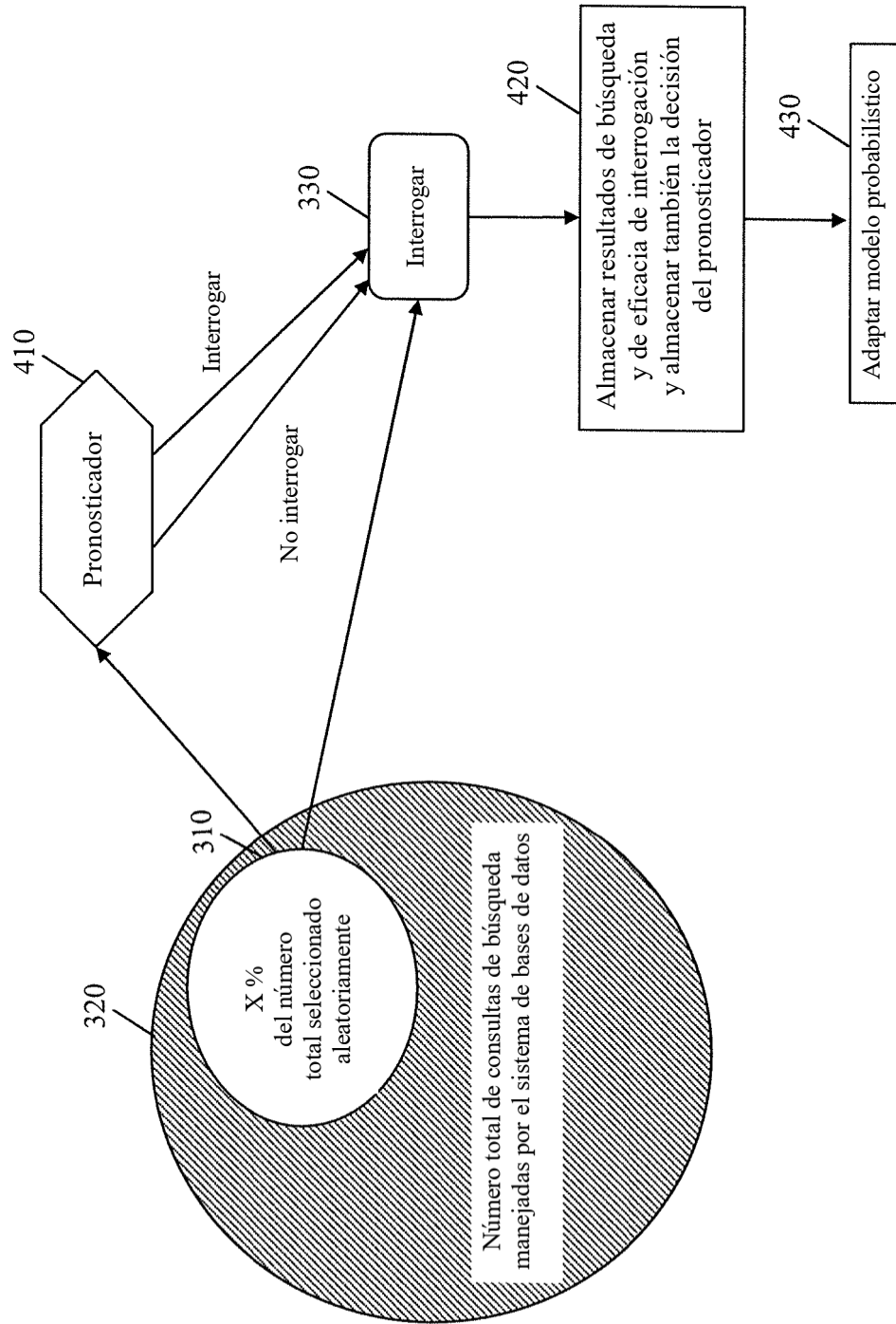


FIG. 4

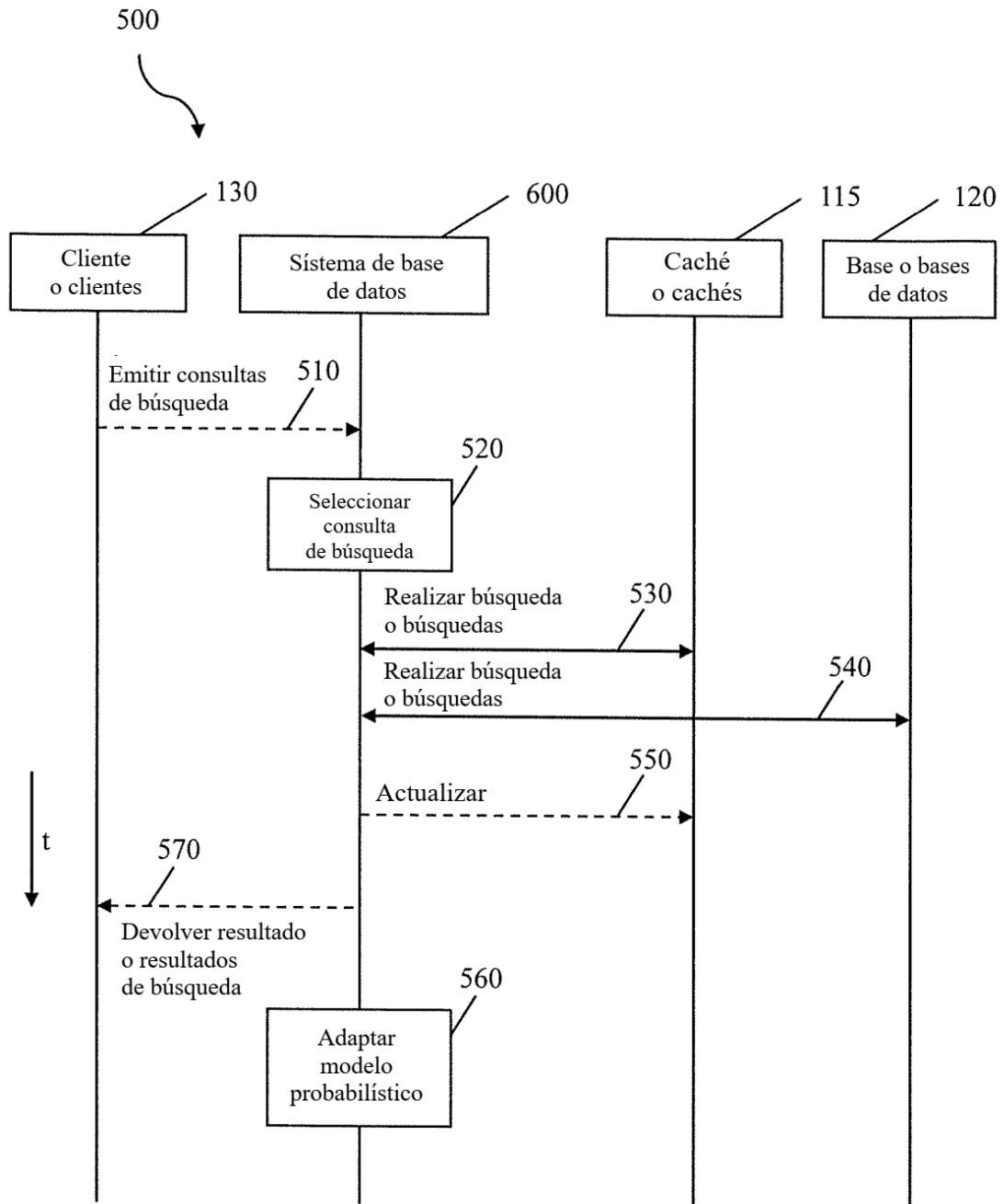


FIG. 5

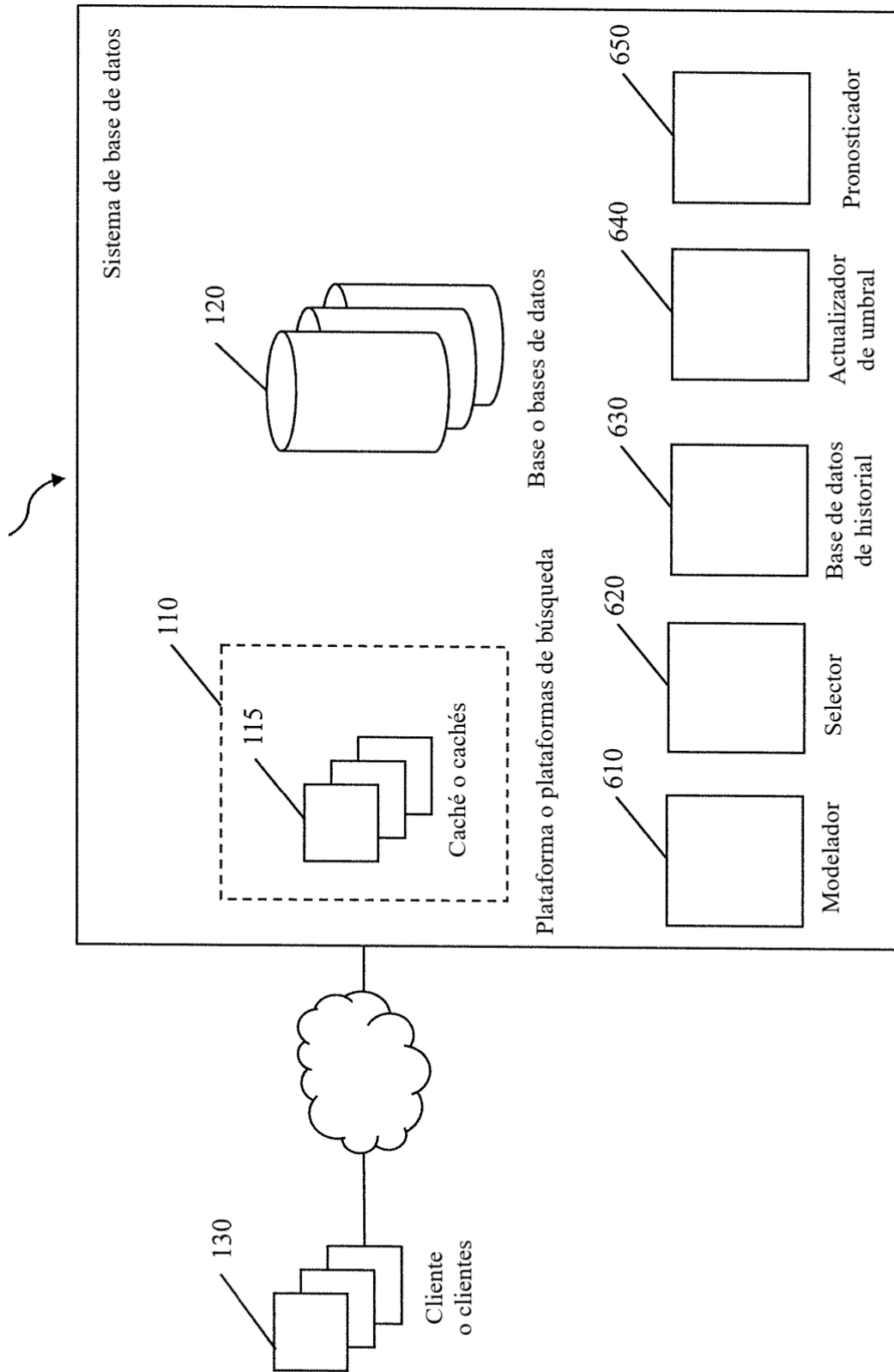


FIG. 6

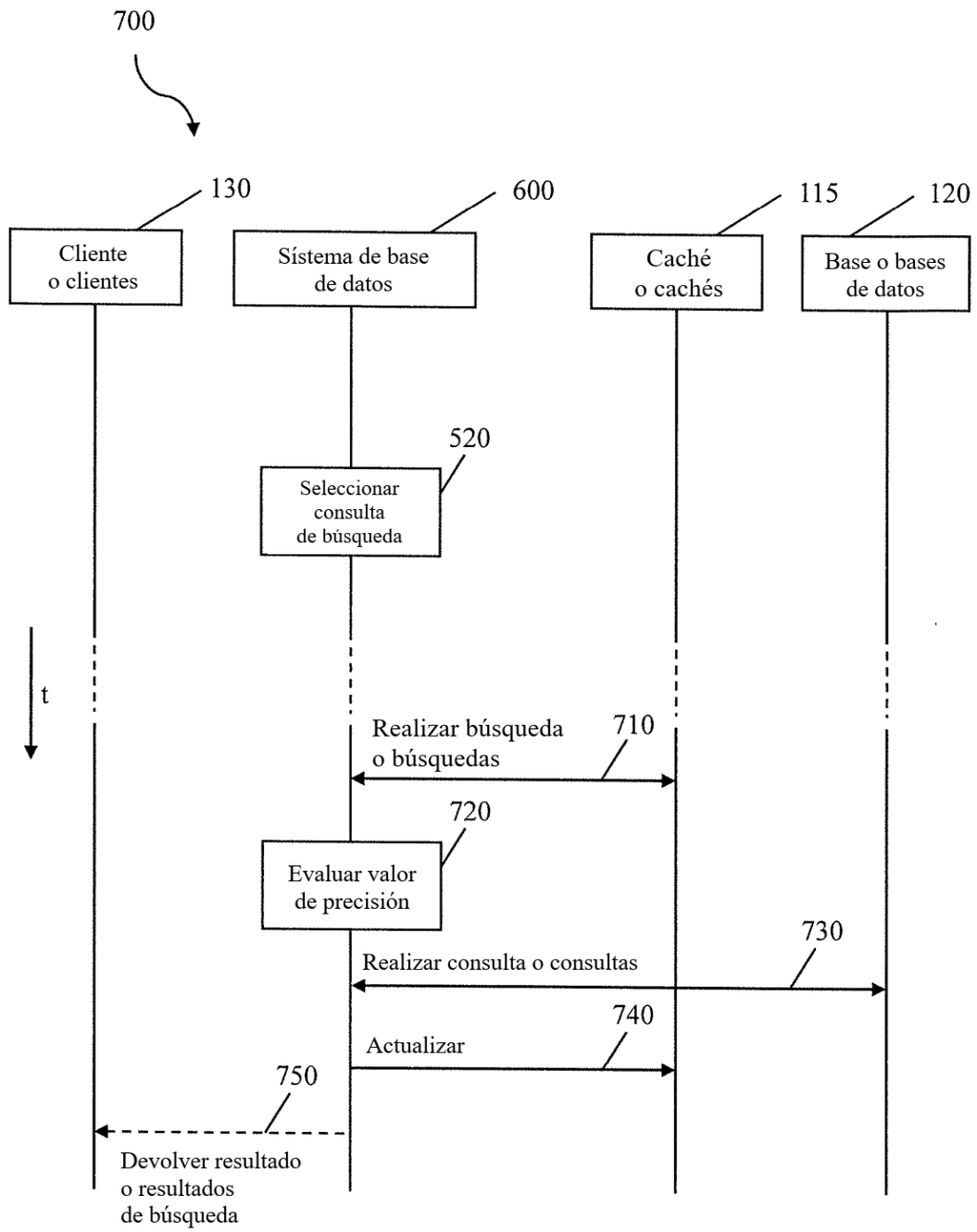


FIG. 7

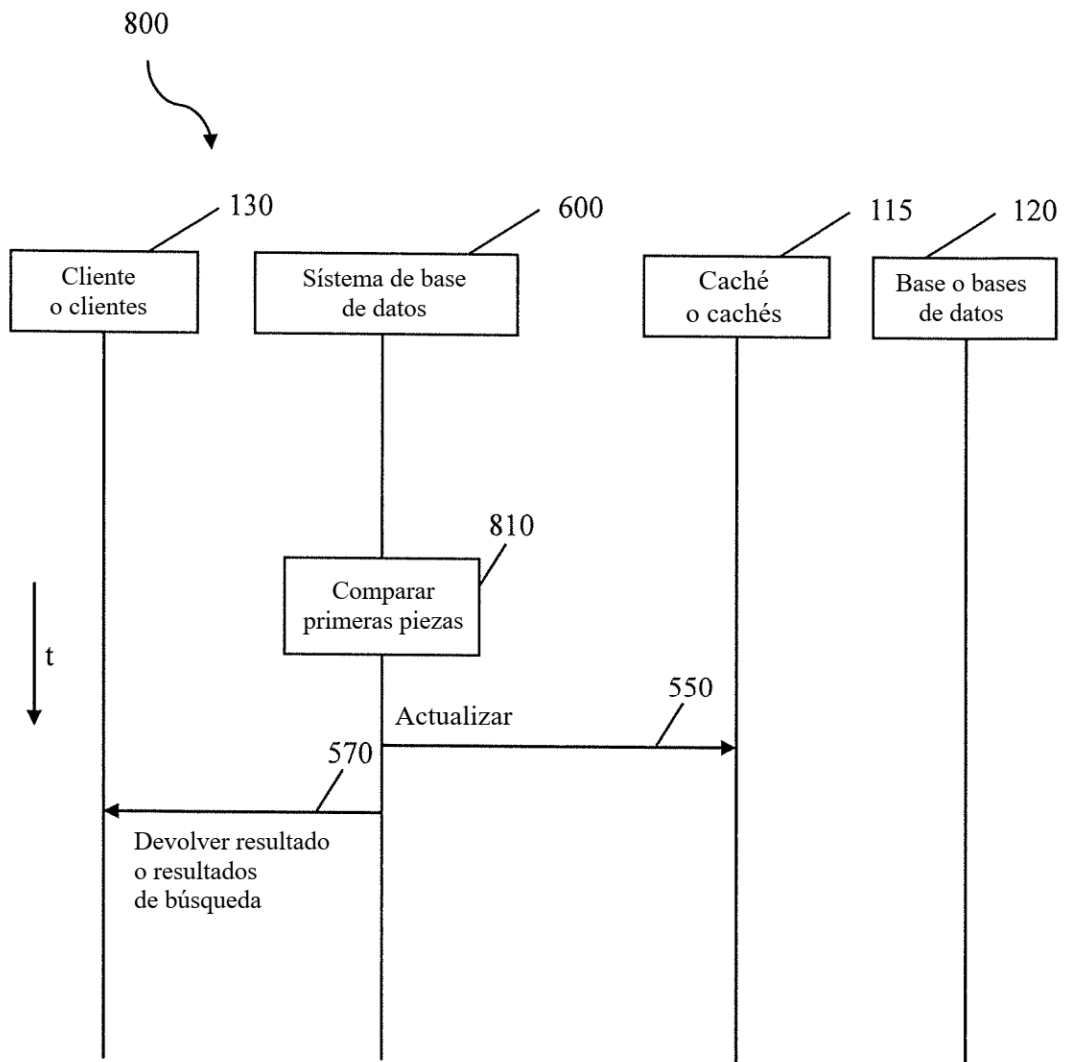


FIG. 8

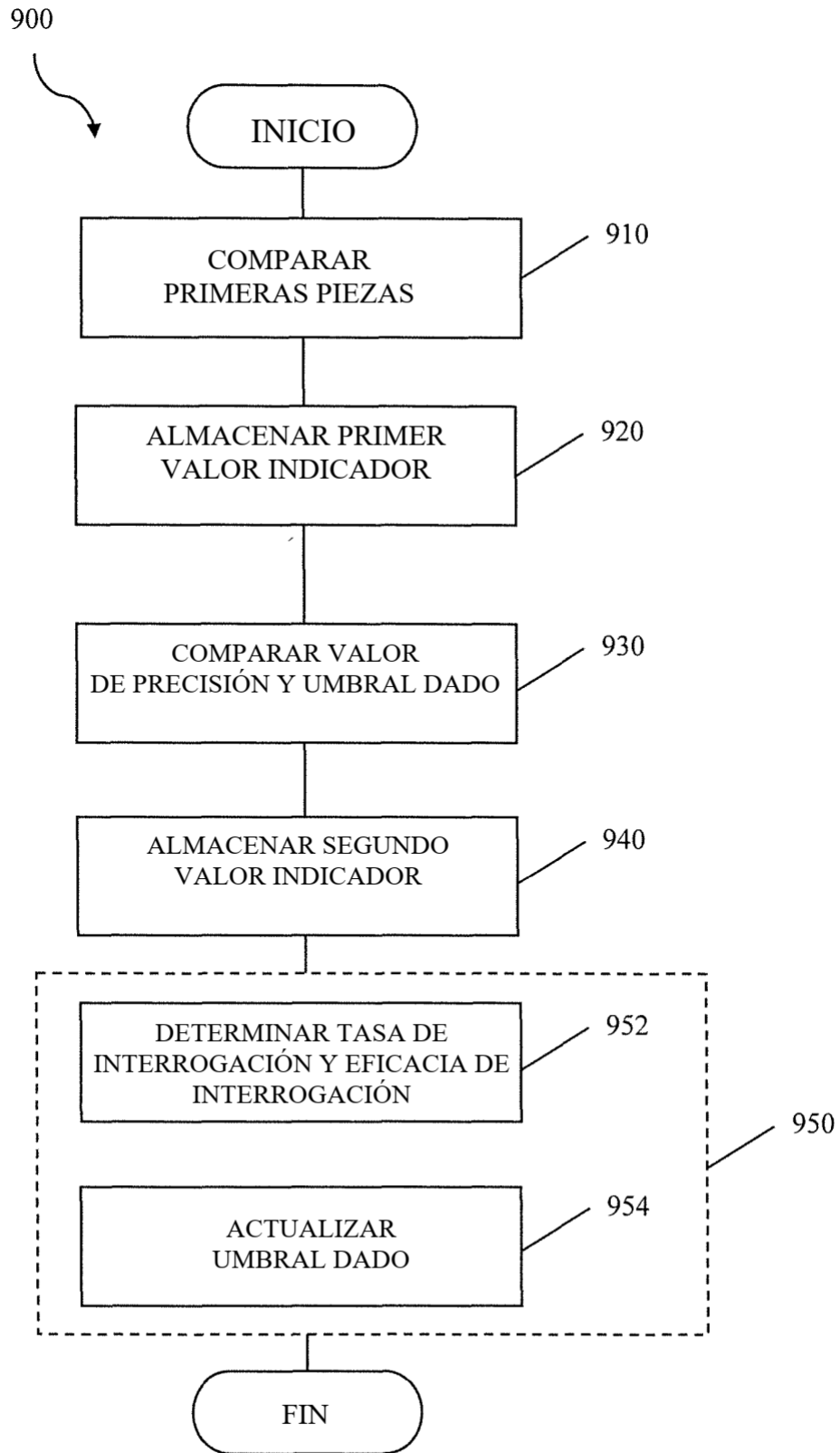


FIG. 9

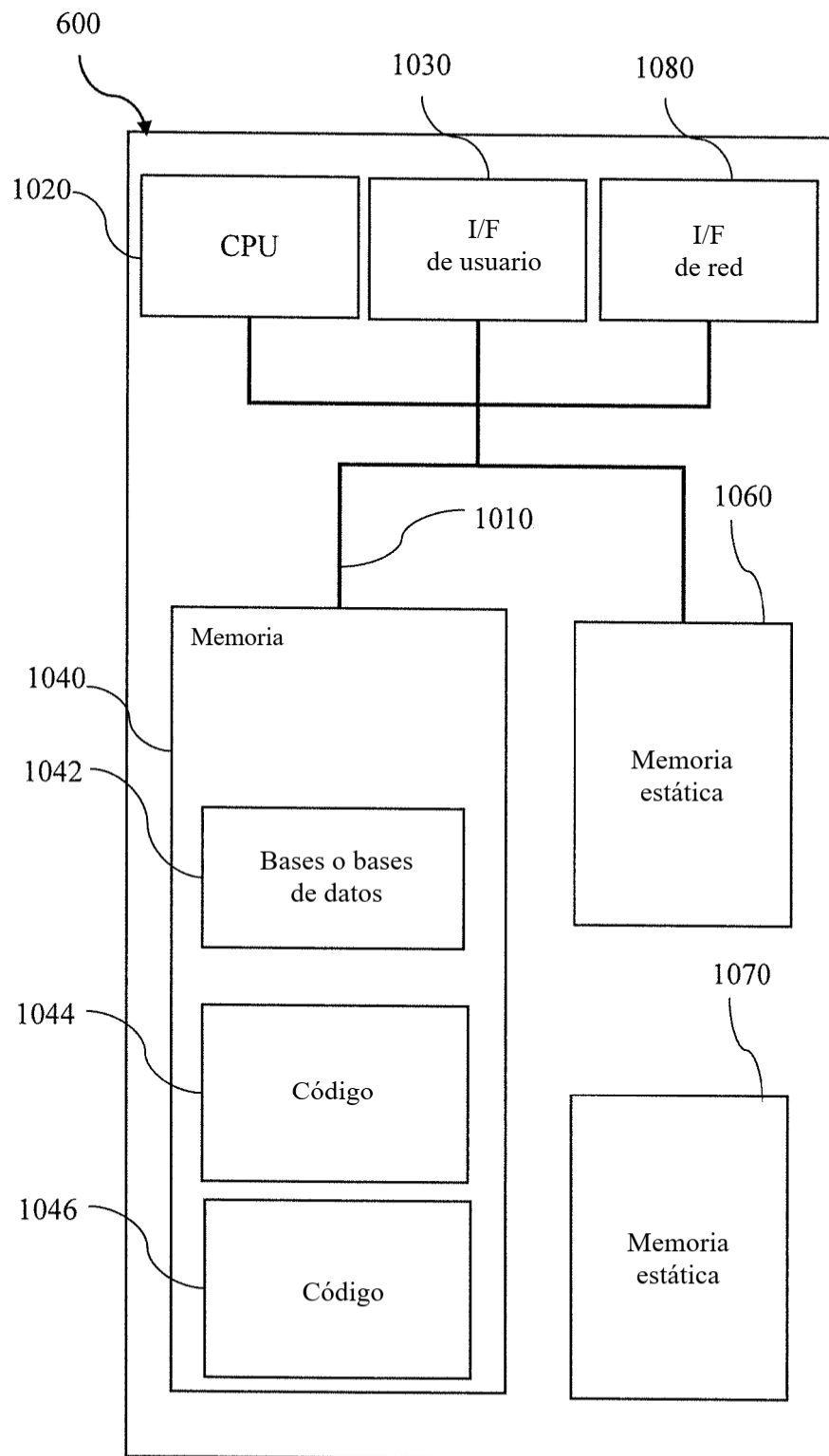


FIG. 10