

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 662 101**

51 Int. Cl.:

G06F 17/30 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.02.2014** **E 14003986 (8)**

97 Fecha y número de publicación de la concesión europea: **13.12.2017** **EP 2913764**

54 Título: **Re-calcular resultados de búsqueda pre-calculados**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.04.2018

73 Titular/es:
AMADEUS S.A.S. (100.0%)
485 Route du Pin Montard Sophia Antipolis
06410 Biot, FR

72 Inventor/es:
LEGRAND, GUILLAUME y
CIABRINI, DAMIEN

74 Agente/Representante:
SUGRAÑES MOLINÉ, Pedro

ES 2 662 101 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

Re-calcular resultados de búsqueda pre-calculados

5 Campo de la invención

La presente invención se refiere en general a tecnología de base de datos. Más específicamente, se refiere a una estrategia para mantener resultados de búsqueda precalculados almacenados como registros de base de datos en una base de datos actualizada.

10

Antecedentes

Un problema común en la tecnología de base de datos es asegurar tiempos de respuesta cortos a consultas de base de datos que requieren procesar grandes volúmenes de datos. Por ejemplo, tal potencia informática que consume procesamiento tiene que realizarse en respuesta a las denominadas "consultas abiertas" que contienen únicamente poca información de entrada (por ejemplo, únicamente se especifican 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, conducen en general a un gran número de resultados. 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.

15

20

Un enfoque general para acortar tiempos de consulta es precalcular consultas esperadas y mantener los resultados de consulta correspondientes en un sistema de caché. Las consultas entonces realmente no se procesan en las bases de datos grandes, sino que se dirigen al sistema de caché.

25

Sin embargo, otro problema que surge junto con tales enfoques de almacenamiento en caché, es mantener los resultados de búsqueda precalculados actualizados para asegurar que las consultas respondidas por los resultados almacenados en caché reflejan correctamente el estado de las correspondientes bases de datos grandes. En el caso de cambios de datos subyacentes, los resultados de consulta almacenados en caché quedan desactualizados y el sistema de caché devolvería resultados incorrectos. Por lo tanto, se emplean estrategias de actualización para mantener el sistema de caché actualizado.

30

Son conocidas en la técnica anterior diversas estrategias de actualización relativamente sencillas como, por ejemplo, recalcular todo el dominio de datos con frecuencia, establecer y mantener planificaciones de recálculo manualmente y recalcular datos cuando se estén volviendo demasiado antiguos.

35

Se han desarrollado estrategias de actualización algo más sofisticadas, como se describen por ejemplo mediante el documento WO 01/33472 y el documento WO 02/25557.

40

El documento WO 01/33472 se refiere a un sistema de disponibilidad usado en un sistema de planificación de viajes. El sistema incluye una caché que tiene entradas de información de disponibilidad con respecto a asientos de la línea aérea. Un gestor de caché gestiona información de entrada en la caché para mantener la información en la caché correcta, actual, completa o de otra manera tan útil como sea posible. En respuesta a una consulta dirigida a la caché, el gestor de caché determina si una respuesta almacenada es obsoleta y, si este es el caso, envía una consulta de disponibilidad a una fuente de información de disponibilidad. Las entradas de caché a modificar se obtienen por notificaciones asíncronas a partir de sistemas externos y se determinan por un modelo determinístico, predictivo o estadístico.

45

De manera similar, el documento WO 02/25557 pertenece a un sistema de recuperación de información en el que la información recibida desde fuentes de información se almacena en caché para uso futuro, tal como para futuras solicitudes de clientes. Las consultas proactivas pueden generarse para rellenar una caché y/o para actualizar información actualmente almacenada en caché. En un sistema de información de línea aérea, las consultas proactivas se ordenan basándose en estadísticas o indicaciones predictivas tales como una cercanía de tiempo de salida, la antigüedad de los datos almacenados en caché, asientos restantes en una aeronave, vacaciones o eventos especiales o tipo de equipaje. Además, se reciben actualizaciones mediante notificaciones externas desde líneas aéreas tales como mensajes AVS.

50

55

Además, el documento WO 99/22315 describe un mecanismo para refrescar automáticamente documentos en una caché usando un modelo probabilístico basado en estadísticas. Para cada documento, la caché determina una probabilidad $Psi(t)$ de que un objeto almacenado en caché i esté obsoleto en un tiempo particular t (es decir el servidor ha cambiado ese objeto) y una probabilidad $Pri(h)$ de que ese objeto i se solicite por un usuario en el tiempo de solicitud h . La caché refresca estos objetos con el producto más alto $Pi = Psi(t) \times Pri(h)$, es decir la probabilidad de que un objeto desactualizado se devuelva al usuario con la siguiente solicitud. Para mantener estos valores de probabilidad, la caché mantiene y registra estadísticas históricas para los objetos almacenados en caché tales como un intervalo medio estimado entre actualizaciones de servidor EUI. El EUI de un objeto se actualiza, por ejemplo, cuando el propio objeto se actualiza por el servidor o el objeto no se actualiza después de que su tiempo de refresco

60

65

medio estimado ha transcurrido.

El documento US 2005/234971 se refiere a refrescar vistas materializadas (MV) en un sistema de base de datos. Se usan costes estimados para planificar el refresco de unas MV establecidas en un orden de manera que se reduce el tiempo global para refrescar, o en un orden que usa el tiempo global más pequeño, para refrescar cuando se compara con otras planificaciones de refresco alternativas. Una ordenación del refresco de las MV determina cuáles MV se refrescan usando otras MV, puesto que el refresco de una MV dada puede realizarse mediante una reescritura de todas las MV que están recientes en el momento que se refresca la MV dada. Las MV se planifican para refrescarse en un orden de manera que las MV usadas por otras MV para refrescar se planifican en primer lugar, y de manera que se minimiza el coste de refrescar las colecciones completas de las MV o al menos es menor que lo que era en las MV refrescadas en uno o más otros posibles órdenes para refrescar las MV. La planificación del orden en el que refrescar las MV se determina planificando recursivamente una MV en un tiempo que se inicia desde aquellas MV que se refrescan directamente desde bases de datos. Se generan expresiones de refresco para las MV que necesitan refrescarse. El documento US 2009/204753 se refiere a refrescar una caché basándose en respuestas de consulta proporcionadas por un sistema de búsqueda en respuesta a consultas. Una entrada de caché se proporciona para cada única consulta si hay espacio disponible en la caché. Se asigna un valor de temperatura a cada entrada de caché basándose en una frecuencia de ocurrencia de la correspondiente consulta. Se asigna un valor de antigüedad a cada entrada de caché basándose en un tiempo de último refresco o creación de la correspondiente respuesta de consulta. La antigüedad de las entradas de caché se actualiza periódicamente y la temperatura de una entrada de caché se actualiza cuando tiene lugar una consulta correspondiente. Si hay recursos de sistema disponibles, la respuesta de consulta de una entrada de caché se refresca basándose en la temperatura y antigüedad de la entrada de caché. Si no hay recursos disponibles, el refresco está limitado

Sumario de la invención

La presente invención se define mediante las reivindicaciones independientes.

De acuerdo con un aspecto, la presente invención proporciona un método de determinar resultados de búsqueda precalculados para recálculo en un entorno de base de datos. El entorno de base de datos está al menos compuesto de al menos una plataforma de búsqueda que mantiene resultados de búsqueda precalculados, un controlador de recálculo y una plataforma de cálculo. El controlador de recálculo genera órdenes de recálculo para recalculados resultados de búsqueda precalculados y transmite las órdenes de recálculo a la plataforma de cálculo. El número de resultados de búsqueda precalculados recalculados por la plataforma de cálculo está limitado por los recursos de cálculo de la plataforma de cálculo disponibles para el recálculo dentro de un intervalo de tiempo dado que forma un ciclo de recálculo.

Los recursos de cálculo para recalculados un resultado de búsqueda precalculado i dependen de si se recalculan o no otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i durante el intervalo de tiempo dado. El controlador de recálculo estima dinámicamente los recursos de cálculo necesarios para recalculados el resultado de búsqueda precalculado i dependiendo de qué otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado se seleccionen para recálculo durante el intervalo de tiempo dado.

Los resultados de búsqueda precalculados para recálculo por la plataforma de cálculo dentro del intervalo de tiempo dado se seleccionan de manera iterativa. Esta selección iterativa incluye una estimación iterativa de los recursos de recálculo para recalculados los resultados de búsqueda precalculados que a su vez comprende:

- a) inicializar los recursos de cálculo c_i necesarios para recalculados cada resultado de búsqueda precalculado i con un valor que asume el cálculo de i independiente del cálculo de otros resultados de búsqueda precalculados seleccionados para recálculo durante el intervalo de tiempo dado;
- b) en una primera iteración, seleccionar una primera porción de los resultados de búsqueda precalculados para recálculo de acuerdo con el indicador de recálculo, en el que - basándose en los valores actuales para los recursos de cálculo necesarios para recalculados la porción de resultados de búsqueda precalculados - la primera porción seleccionada únicamente requiere un porcentaje por debajo del 100 % de los recursos de cálculo totales disponibles de la plataforma de cálculo para recálculo en el intervalo de tiempo dado;
- c) reevaluar a los recursos de cálculo c_i necesarios para recalculados cada resultado de búsqueda precalculado i teniendo en cuenta qué resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i se han seleccionado para recálculo;
- d) realizar iteraciones adicionales de selección de porciones aún no seleccionadas adicionales de resultados de búsqueda precalculados y reevaluar los recursos de cálculo c_i necesarios para recalculados cada resultado de búsqueda precalculado i si se agota menos del 100 % de los recursos de cálculo globales disponibles de la plataforma de cálculo para recálculo en el intervalo de tiempo dado.

Además, se proporciona un controlador de recálculo y un programa informático que están configurados para realizar estas actividades.

Se definen aspectos adicionales mediante las reivindicaciones dependientes.

Breve descripción de las figuras

5 La presente invención se describirá con referencia a las figuras adjuntas. Números de referencia similares indican en general elementos idénticos o de funcionalidad similar.

La Figura 1 muestra esquemáticamente un entorno de base de datos distribuido.

10 La Figura 2 ilustra un modelo probabilístico que predice precisión decreciente de un resultado de búsqueda precalculado durante el tiempo.

Las Figuras 3a, 3b y 3c visualizan los efectos de una estrategia de recálculo que se centra en recalcular resultados de búsqueda precalculados volátiles.

15 Las Figuras 4a y 4b visualizan los efectos de una estrategia de recálculo teniendo en cuenta frecuencias de recálculo, costes de recálculo para recalcular resultados de búsqueda precalculados y su popularidad de búsqueda.

La Figura 5 representa costes de recálculo para diferentes tipos de conjuntos de resultados de búsqueda precalculados.

20 La Figura 6 muestra un ejemplo de un entorno de base de datos que implementa los métodos presentados en el presente documento.

La Figura 7 presenta una vista en una estructura interna ejemplar del controlador de recálculo.

La Figura 8 es una vista esquemática ejemplar de la arquitectura interna del servidor de procesamiento de consulta.

Descripción detallada

25 Para poder manejar consultas de base de datos o solicitudes de cálculo en lotes que requieren cálculos basándose en grandes volúmenes de datos subyacentes, los resultados de búsqueda que corresponden a consultas esperadas generalmente se precalculan y almacenan como registros de base de datos en una base de datos. Se consulta esta base de datos por entidades solicitantes (tales como clientes, aplicaciones, exploradores instalados en terminales de usuario, etc.) en el curso de una búsqueda y se devuelven resultados de búsqueda precalculados que satisfacen criterios de búsqueda indicados por la consulta al cliente en respuesta a la consulta. Posteriormente, el término "consulta" se usa como un término general que incluye cualquier tipo de solicitudes de recuperación de información tales como consultas transaccionales, peticiones para cálculos en lotes y otras formas.

35 La Figura 1 ilustra un entorno de base de datos 1 de este tipo en un nivel abstracto. Los datos básicos, en lo sucesivo denominados también como "datos de cálculo", se mantienen en y/o se acceden por una plataforma de cálculo 3 que está conectada a un controlador de recálculo 2. El último emite órdenes de recálculo a la plataforma de cálculo 3 que, a su vez, transmite los resultados correspondientes a la plataforma de búsqueda 4 y, además, al controlador de recálculo 2 que también mantiene los resultados de búsqueda precalculados por razones de control de recálculo. Los usuarios finales 5 tales como aplicaciones en terminales de usuario acceden a los resultados de búsqueda precalculados a partir de la plataforma de búsqueda 4. Como se indica mediante la Figura 1, una o varias plataformas de búsqueda 4 pueden estar presentes en el entorno 1. Si está presente una pluralidad de plataformas de búsqueda 4, los resultados de búsqueda precalculados pueden mantenerse de una manera distribuida a través de las varias plataformas de búsqueda 4 y el controlador de recálculo 2 puede controlar el recálculo de todos los resultados de búsqueda precalculados distribuidos a través de las plataformas de búsqueda 4. La plataforma de búsqueda 4 puede ofrecer también un conjunto heterogéneo de solicitudes de búsqueda precalculadas, por ejemplo algunas plataformas de búsqueda 4 mantienen solicitudes de búsqueda precalculadas relacionadas con un viaje aéreo, otras plataformas de búsqueda 4 almacenan solicitudes de búsqueda precalculadas relacionadas con seguros y otras plataformas de búsqueda 4 mantienen solicitudes de búsqueda precalculadas (o pre-rastreadas) relacionadas con sitios web de internet. Un entorno heterogéneo de este tipo puede controlarse por un único controlador de recálculo 2 o por una pluralidad de controladores de recálculo 2. Como alternativa, la pluralidad de plataformas de búsqueda 4 puede utilizarse para espejar los mismos resultados de búsqueda precalculados, por ejemplo, por razones de redundancia.

55 El enfoque de precalcular resultados de búsqueda y almacenarlos en la plataforma de búsqueda accesible a clientes de consulta conduce a la situación general de que los datos de cálculo pueden cambiar durante el tiempo y, por lo tanto, los resultados de búsqueda precalculados quedan desactualizados o inválidos (ambos términos se usan de manera sinónima en el presente documento). Los resultados de búsqueda precalculados que aún están actualizados, es decir que coinciden con los equivalentes de cálculo en tiempo real correspondientes (resultados que se calcularían realmente bajo demanda sin tener resultados de búsqueda precalculados disponibles), se denominan resultados de búsqueda precalculados "precisos" en lo sucesivo. Por lo tanto, cuando la plataforma de búsqueda que mantiene los resultados de búsqueda precalculados representa correctamente el estado actual del dominio de datos que subyace los resultados de consulta almacenados en caché, es decir los datos de cálculo, los resultados de búsqueda precalculados almacenados en la plataforma de búsqueda son, en general, precisos.

65

En general, para posibilitar que la plataforma de búsqueda devuelva resultados correctos, se desea mantener un alto grado de correlación entre resultados de búsqueda precalculados que se proporcionan a la entidad de consulta en respuesta a consultas de base de datos y sus equivalentes de cálculo en tiempo real. Al mismo tiempo, sin embargo, es deseable minimizar el consumo de recursos de cálculo producido por recálculos, es decir evitar cualquier recálculo innecesario tal como recálculo de resultados de consulta almacenados en caché ya precisos. Los recursos informáticos son limitados y, en general, no hay suficientes recursos informáticos para recalcular todos los resultados de búsqueda precalculados en todo momento. Por lo tanto, ha de hallarse un equilibrio entre precisión de resultados de búsqueda precalculados y utilización de la potencia informática disponible.

Para hacer el recálculo más eficaz, se definen métricas para evaluar cómo es de “necesario” o “innecesario” un recálculo. Por ejemplo, no merece la pena relanzar un pre-cálculo masivo completo cada día si menos de la mitad de los resultados de consulta calculados se vuelven desactualizados. Por otra parte, si se sabe que clases particulares de resultados de consulta cambian frecuentemente, recalcularlas varias veces por día puede ser beneficioso para la precisión. En consecuencia, es necesaria una manera efectiva de evaluar o estimar precisión de resultado de búsqueda, teniendo en cuenta en general tanto la ganancia asociada en la precisión como el coste de recálculo.

De acuerdo con la estrategia de actualización de resultados de búsqueda presentada en el presente documento, se deciden los recálculos de resultados de búsqueda precalculados basándose en un modelo predictivo que produce estimaciones de la precisión de los resultados de búsqueda precalculados mantenidos en la plataforma de búsqueda. Este modelo predictivo modela las discrepancias entre los resultados de búsqueda precalculados y resultados de búsqueda reales supuestos, es decir aproxima la precisión o imprecisión de cualquier resultado de búsqueda precalculado. El modelo modela, por ejemplo, la validez probable de los resultados de búsqueda precalculados durante el tiempo. Las suposiciones sobre la validez de los resultados precalculados se concluyen y extrapolan a partir de las experiencias del mundo real pasadas en la materia objeto del respectivo dominio de datos.

Por ejemplo, los datos de cálculo subyacente pueden pertenecer al dominio de viajes aéreos y contener información sobre vuelos tales como aeropuerto de salida y destino, línea aérea, fechas de salida y retorno, tarifas, clases de reserva y similares. Estos datos relacionados con viajes aéreos se mantienen en la plataforma de cálculo y se consultan por clientes para tener conocimiento de, por ejemplo, disponibilidad de precios de vuelos aéreos o cualquier otro producto/servicio de viaje con precio. Por ejemplo, calcular precios basándose en los datos de vuelo básicos consume recursos y tiempo. Por lo tanto, los precios reales se precalan y almacenan en la plataforma de búsqueda. En este ejemplo, el modelo probabilístico modela la validez de los precios de recomendación de viaje durante el tiempo. El conocimiento requerido para crear un modelo de este tipo puede tomarse a partir de las experiencias del mundo real sobre el comportamiento y desarrollo de, por ejemplo, precios de recomendación de viaje anteriores a la fecha de salida. Por ejemplo, puede conocerse que los precios permanecen relativamente estables durante el periodo de tiempo anterior a un mes antes de las respectivas fechas de salida, pero se vuelven más volátiles durante el mes antes de la fecha de salida. Por lo tanto, el modelo probabilístico indica que los precios precalculados que pertenecen a recomendaciones de viaje próximas en el siguiente mes deberían recalcularse más a menudo que tales precios precalculados que están asociados con las recomendaciones en el futuro más distante.

Una estrategia particular para recalcular resultados de búsqueda precalculados basándose en un modelo predictivo se propone, por ejemplo, por la solicitud de patente internacional no publicada PCT/EP2013/002390 que se incluye por referencia en el presente documento. En particular, los enfoques desvelados en la misma tienen en cuenta la denominada “volatilidad” y “popularidad” de los resultados de búsqueda precalculados. Obsérvese que esta “volatilidad” no se refiere a un cambio de los mismos resultados de búsqueda precalculados, sino al cambio de los datos subyacentes que forman la base para el cálculo de los resultados de búsqueda precalculados, dando como resultado potencialmente resultados de búsqueda precalculados desactualizados o inválidos. La “popularidad” se refiere a la frecuencia de consultas a los resultados de búsqueda precalculados en la base de datos: los resultados de búsqueda precalculados se consultan más a menudo que otros que son más populares que estos otros resultados de búsqueda precalculados menos populares.

Los presentes inventores han reconocido que concentrar la estrategia de recálculo únicamente en resultados de búsqueda precalculados volátiles y populares (es decir tales resultados de búsqueda precalculados que se desactualizan regularmente por cambio de sus correspondientes datos de cálculo subyacentes) no es óptimo. Los inventores se han dado cuenta de que tal estrategia únicamente produce aumentos a corto plazo de la precisión de los resultados de búsqueda precalculados, pero puede provocar una degradación de la precisión a largo plazo. Esto se describe ahora en más detalle en primer lugar, antes de volver a las consecuencias de estos reconocimientos extraídos por la presente invención.

Para entender las desventajas de una estrategia de recálculo que se centra de manera particular únicamente en la volatilidad y popularidad de resultados de búsqueda precalculados, se definen los siguientes parámetros para un resultado de búsqueda precalculado i:

- La antigüedad t_i del resultado de búsqueda precalculado i: el tiempo desde el último cálculo de este resultado de búsqueda precalculado por la plataforma de cálculo 3.

- La tasa de validez λ_i del resultado de búsqueda precalculado i es una medida de cuánto tiempo el resultado de búsqueda precalculado i permanece válido o con qué rapidez el resultado de búsqueda precalculado i se vuelve inválido debido a cambios de los datos originales subyacentes. Esta tasa de validez de un resultado de búsqueda precalculado dado i se deriva estadísticamente, por ejemplo, a partir de las apariciones y los resultados (re-)cálculos o (re-)cogidas y comparaciones del pasado del resultado de búsqueda recogido con su estado o valores anteriores. Por ejemplo, se ha determinado que un resultado de búsqueda pre-recogido particular i tiene una tasa de validez λ_i del 10 % por hora, que significa que la probabilidad de que i sea válido se reduce un 10 % cada hora. En el momento de su (re-)cogida o (re-)cálculo, i es en general el 100 % válido. Después de una hora, i es válido con una probabilidad del 90 %. Después de dos horas la validez de i es del 81 % (=90 % reducido por otro 10 %). Después de tres horas, la validez probable de i está al 72,9 %, y así sucesivamente.

- La tasa de validez λ_i puede emplearse para proporcionar una estimación de la probabilidad de que un resultado de búsqueda precalculado permanezca válido después de un tiempo dado: **$P(\text{invariable después de } t) = e^{-\lambda_i t}$** . Esto se refiere también a la precisión esperada **$acc_i^t = e^{-\lambda_i t}$** o, más en general, como la probabilidad de que un resultado de búsqueda precalculado sea válido o, en otras palabras, no esté desactualizado. Se representan dos funciones ejemplares de esta reducción de precisión probable por la Figura 2. La función 10 representa un resultado de búsqueda precalculado que potencialmente permanece más preciso (o, más correctamente, permanece a una probabilidad superior de ser válido durante el tiempo) que otro resultado de búsqueda precalculado asociado con la función 11. Por ejemplo, el resultado de búsqueda precalculado representado por la función 10 tiene el 70 % de probabilidad de que aún sea válido a 35 horas después de su último recálculo, mientras que el otro resultado de búsqueda precalculado caracterizado por la función 11 es únicamente válido hasta aproximadamente el 50 % a las 35 horas después de su último recálculo. Las funciones 10 y 11 pueden representar también conjuntos completos de resultados de búsqueda precalculados y a continuación indicar proporciones de los conjuntos de resultados de búsqueda precalculados que es probable que sean válidos en un tiempo pasado desde el último recálculo del conjunto.

- La precisión esperada inicial: resultados de búsqueda precalculados generados por la plataforma de cálculo 3 pueden no ser necesariamente precisos incluso en el tiempo de cálculo si, por ejemplo, la misma plataforma de cálculo 3 basa sus cálculos en datos almacenados en caché (y por lo tanto desactualizados). Esto conduce a discrepancias adicionales entre resultados de búsqueda precalculados calculados por la plataforma de cálculo 3 y resultados de cálculo que se han generado hipotéticamente por datos subyacentes precisos. Esta discrepancia puede medirse si está disponible la respectiva realimentación. Puede inferirse, por ejemplo, a partir de los cálculos anteriores que el resultado de búsqueda precalculado i tiene una probabilidad a_i de que sea preciso en el tiempo de cálculo por la plataforma de cálculo 3. Significa que la probabilidad para que un resultado de búsqueda precalculado sea preciso después de un tiempo dado t es **$a_i e^{-\lambda_i t}$** .

- La precisión de los resultados de búsqueda precalculados totales mantenidos en la plataforma de búsqueda 4 de acuerdo con este modelo ejemplar puede considerarse entonces como la precisión media ("precisión global"):

$$\text{Precisión Global} = \text{Avg} (a_i e^{-\lambda_i t_i}).$$

- La "popularidad" p_i del resultado de búsqueda precalculado i : esto es la frecuencia de acceso media a este resultado de búsqueda precalculado por los usuarios finales. Como ya se ha indicado brevemente antes, puede ser deseable conseguir una mejor precisión para estos resultados de búsqueda precalculados que se solicitan más a menudo por los usuarios que otros resultados de búsqueda precalculados. Por lo tanto, la precisión de la suma total de los resultados de búsqueda precalculados en la base de datos 2 como se observa por los usuarios finales puede definirse también en que cada valor de precisión se pondera por la popularidad del respectivo resultado de búsqueda precalculado. Por lo tanto, la proporción de los accesos precisos a los resultados de búsqueda precalculados a diferencia de la proporción esperada de los resultados de búsqueda precalculados precisos está incluida:

$$\text{Precisión de Usuario} = \sum \frac{p_i}{p_{tot}} a_i e^{-\lambda_i t_i} \text{ en la que } p_{tot} \text{ se define como la suma de toda } p_i.$$

Estos parámetros asignados a cualquier resultado de búsqueda precalculado a partir de un modelo de precisión probabilístico utilizado para decidir qué resultados de búsqueda precalculados en la base de datos 4 reciben prioridad para recálculo. Los valores de estos parámetros se derivan a partir de y se establecen, por ejemplo, de acuerdo con datos de historial estadístico derivados de un servidor de estadísticas basándose en varios días, semanas o meses, etc., de historial. Cada único resultado de búsqueda precalculado se modela con estos parámetros para predecir el estado de los resultados de búsqueda precalculados y por lo tanto predecir la precisión de todos los resultados de búsqueda precalculados.

En general, existen varias maneras para utilizar esta información del modelo probabilístico para priorizar y decidir qué resultados de búsqueda precalculados recalculan a continuación. Estos resultados de búsqueda precalculados priorizados se denominan también como los “más críticos” de recalculan. Un enfoque natural para crear un criterio de criticidad sería hallar los órdenes de cálculo que se asumen que aumentan como máximo la precisión del conjunto global de precalcular solicitudes de búsqueda almacenadas en la base de datos 4 desde la perspectiva del usuario final (“Precisión de usuario”), es decir estas solicitudes de búsqueda precalculadas con la probabilidad más alta de que sean inválidas, ponderadas por su popularidad. Una estrategia de este tipo se desvela, por ejemplo, por el documento WO 99/22315 proponiendo refrescar estos objetos en una caché con el producto más alto $P_i = \Psi(t) \times \text{Pri}(h)$.

El enfoque determina el conjunto de resultados de búsqueda precalculados C a recalculan que:

- aumenta como máximo la precisión del usuario del conjunto total de resultados de búsqueda precalculados almacenados en la base de datos 4 (brevemente “dd” para el “dominio de datos”):

$$\text{Precisión de Usuario} = \sum_{x_i \in \text{dd}} \frac{P_i}{P_{\text{tot}}} \alpha_i e^{-\lambda_i t_i}$$

- y
- cuesta menos que R , indicando R los recursos disponibles para el siguiente ciclo de recálculo:

$$\text{Coste}(C) = \sum_{x_i \in C} c_i \leq R$$

Recalculan el resultado de búsqueda precalculado i aumentaría la precisión de usuario por $\alpha_i \frac{P_i}{P_{\text{tot}}} (1 - e^{-\lambda_i t_i})$,

que se denomina como ganancia $G_i = \alpha_i \frac{P_i}{P_{\text{tot}}} (1 - e^{-\lambda_i t_i})$.

Este enfoque podría perfeccionarse adicionalmente considerando los costes de cálculo (es decir recursos informáticos) requeridos para recalculan un conjunto de solicitudes de búsqueda precalculadas. Si los costes de cálculo para recalculan la solicitud de búsqueda precalculada x_i se indican como c_i , se busca la relación ganancia-coste más alta para aumentar la “precisión de usuario” más eficazmente, es decir mejor aumento con mínima utilización de recursos de cálculo. Esta relación ganancia-coste puede definirse para que sea

$$\frac{G_i}{c_i} = \alpha_i \frac{P_i}{P_{\text{tot}}} \frac{1 - e^{-\lambda_i t_i}}{c_i}$$

El proceso para determinar los resultados de búsqueda precalculados para que se recalculen a continuación incluye las siguientes dos actividades:

- El controlador de recálculo 2 clasifica la solicitud de búsqueda precalculada por esta relación ganancia-coste.
- El controlador de recálculo 2 selecciona las mejores solicitudes de búsqueda precalculadas a partir de esta lista clasificada hasta que sus costes de cálculo acumulados alcanzan la cantidad de recursos de cálculo R disponibles en la plataforma de cálculo 3, por ejemplo para un cierto periodo de tiempo formando un ciclo de recálculo.

Un comportamiento de la precisión de los resultados de búsqueda precalculados durante el tiempo cuando se emplea este enfoque de recálculo se indica por las Figuras 3a, 3b y 3c. Las tres figuras, así como las Figuras 4a y 4b adicionales, muestran gráficos obtenidos por simulación sobre resultados de búsqueda precalculados artificiales usando una plataforma de cálculo 3 equipada con una cantidad adecuada de recursos de cálculo. Se ha supuesto una fluctuación normal dentro de los datos de cálculo subyacentes, por ejemplo una tasa de validez media aproximadamente del 2 % por hora que significa que el 2 % de los resultados de búsqueda precalculados simulados se quedan desactualizados cada hora (por ejemplo 98 de 100 resultados de búsqueda precalculados son aún válidos después de una hora, después de 10 horas, únicamente $0,98^{10} = 82$ de los 100 resultados de búsqueda precalculados son aún válidos, y así sucesivamente). Debería observarse que este ejemplo de una tasa de validez del 2 % media no implica una tasa de validez homogénea para todos los resultados de búsqueda precalculados simulados. En su lugar, la tasa de validez de resultados de búsqueda precalculados simulados individual puede variar (como lo hace en realidad) y es únicamente de media alrededor del 2 %. Para destacar el desarrollo de la precisión de los resultados de búsqueda precalculados, las simulaciones empiezan desde una situación ideal correspondiendo todos los resultados de búsqueda precalculados a sus datos de cálculo subyacentes, es decir todos los resultados de búsqueda precalculados son válidos y precisos, respectivamente.

La Figura 3a muestra el desarrollo a largo plazo de la precisión global de (todos) los resultados de búsqueda precalculados almacenados en la plataforma de búsqueda 4. Como se ha señalado anteriormente, la expresión “precisión global” se refiere a la popularidad media de todos los resultados de búsqueda precalculados simulados. El número de ciclos de recálculo, que duran por ejemplo 20 minutos cada uno, se expone en el eje X. El porcentaje de resultados de búsqueda precalculados válidos se representa en el eje Y. La función 20 muestra el desarrollo de la validez de los resultados de búsqueda precalculados cuando se emplea una estrategia de recálculo sin tratamiento previo, en concreto recalculando siempre aquellos resultados de búsqueda precalculados que no se han recalculado para el periodo de tiempo más largo (los denominados resultados de búsqueda precalculados “más antiguos”). Puede observarse que la precisión global de los resultados de búsqueda precalculados se reduce rápidamente con las primeras decenas de ciclos de recálculo. Cae por debajo del 75 % alrededor del ciclo de recálculo de orden 45 y a continuación permanece de manera constante entre el 70 % y el 75 %. En comparación con este enfoque de recálculo sin tratamiento previo, la estrategia basándose en la relación ganancia-coste como se ha deducido en detalle anteriormente se muestra mediante el gráfico 21 en la Figura 3a. La estrategia orientada a relación coste-ganancia mostrada mediante la Figura 3a, no considera, sin embargo, la popularidad de los resultados de búsqueda precalculados, es decir la estrategia de recálculo se basa, por ejemplo, en la fórmula $\frac{1 - e^{-\lambda_i t_i}}{c_i}$. Es fácilmente evidente que el enfoque de relación ganancia-coste no mejora significativamente la precisión global de los resultados de búsqueda precalculados. Durante las primeras decenas de ciclos de recálculo, la precisión de los resultados de búsqueda precalculados aún permanece por encima de la función 20, aunque de manera insignificante. Sin embargo, a partir de aproximadamente el ciclo de recálculo 70, la función 21 incluso cae por debajo de la función 20 y permanece únicamente ligeramente por encima del 70 % de precisión a través de todos los siguientes ciclos de recálculo. Por lo tanto, a largo plazo, el enfoque de relación ganancia-coste realmente conduce a una validez reducida de los resultados de búsqueda precalculados, en comparación con la estrategia de recálculo sin tratamiento previo de recalculando siempre los resultados de búsqueda precalculados “más antiguos”.

La Figura 3b muestra los efectos de la estrategia orientada a relación coste-ganancia en relación con la estrategia de recálculo sin tratamiento previo si la estrategia orientada a relación coste-ganancia también considera la “popularidad” de los resultados de búsqueda precalculados, es decir los resultados de búsqueda precalculados más populares se recalculan más a menudo que los resultados de búsqueda precalculados menos populares (que se refleja, como se ha descrito anteriormente, mediante la expresión “precisión de usuario” que implica que la “experiencia del usuario” se mejora por lo que los resultados de búsqueda precalculados que se solicitan más a menudo que otros son más importantes). Por lo tanto, la estrategia orientada a relación coste-ganancia de la Figura 3b está caracterizada, por ejemplo, por la fórmula $p_i \frac{1 - e^{-\lambda_i t_i}}{c_i}$. Como se indica mediante el gráfico 23 de la Figura 3b, la estrategia de recálculo orientada a relación ganancia-coste consigue en general una precisión mejorada de los resultados de búsqueda precalculados que el recálculo sin tratamiento previo (que se muestra mediante el gráfico 22). Esta mejora, sin embargo, se provoca por el hecho de que la estrategia de recálculo sin tratamiento previo no tiene en cuenta, por definición, la popularidad de los resultados de búsqueda precalculados. Esto da como resultado la “subida y bajada” del gráfico 22 debido a que recalculando el resultado de búsqueda no popular conduce a una reducción de la “precisión de usuario”. Por lo tanto, la comparación entre la estrategia sin tratamiento previo y la estrategia orientada a relación coste-ganancia que considera la popularidad se desvía hasta cierto punto. Sin embargo, como mostraremos adicionalmente a continuación, puede extraerse a partir de la Figura 3b (en comparación con la Figura 4a adicionalmente a continuación) que la estrategia orientada a relación coste-ganancia que incluye la popularidad como se visualiza mediante la Figura 3b produce una precisión de usuario que está aún en el intervalo general de la precisión de usuario resultante de la estrategia sin tratamiento previo.

Pueden extraerse conclusiones similares a partir de la Figura 3c que muestra una distribución de precisión acumulativa de todos los resultados de búsqueda precalculados almacenados en la plataforma de búsqueda 4 en el final de la simulación. El eje X indica el porcentaje de resultados de búsqueda precalculados, mientras que el eje Y indica de nuevo el porcentaje de resultados de búsqueda precalculados que es válido. De nuevo, la línea 24 indica la estrategia “recalculando el más antiguo en primer lugar” sin tratamiento previo, mientras que el gráfico 25 muestra la estrategia basándose en la relación ganancia-coste (es decir un valor de 40 % en el eje X y 0,7 en el eje Y indica que el 40 % de los resultados de búsqueda precalculados tienen una precisión de menos del 70 %). Puede tomarse a partir de la Figura 3c que la estrategia de recálculo orientada a relación ganancia-coste consigue algunas mejoras (en comparación con la estrategia sin tratamiento previo) para la mayoría de los resultados de búsqueda precalculados imprecisos. Con esa estrategia, los más imprecisos alrededor del 18 % de todos los resultados de búsqueda precalculados aún tienen una precisión alrededor del 50 %, mientras que la estrategia sin tratamiento previo conduce a significativamente los resultados de búsqueda precalculados más imprecisos del 18 %, aproximadamente el 10 % de los resultados de búsqueda precalculados incluso tienen una precisión por debajo del 40 %. Por otra parte, la estrategia de recálculo sin tratamiento previo parece producir resultados superiores para el otro 80 % más preciso de los resultados de búsqueda precalculados, como puede inferirse del progreso de las curvas 24 y 25 al lado de la derecha de la Figura 3c.

Por lo tanto, las Figuras 3a, 3b y 3c transportan la idea de que la estrategia de recálculo orientada a relación ganancia-coste no es óptima. Al contrario, da como resultado en general una precisión media reducida de resultados de búsqueda precalculados, en comparación con la estrategia sin tratamiento previo para recalcular los resultados de búsqueda precalculados “más antiguos”, cuando se considera el desarrollo a largo plazo. Por lo tanto, seleccionar resultados de búsqueda precalculados para recálculo empleando la relación ganancia-coste aparentemente constituye una optimización a corto plazo únicamente. Los inventores se han dado cuenta que esta estrategia gasta demasiados recursos informáticos en los resultados de búsqueda precalculados muy volátiles, es decir los resultados de búsqueda precalculados que se hacen inválidos más a menudo que otros. Es el objetivo de esta estrategia mantener estos resultados de búsqueda precalculados volátiles a un nivel de precisión aceptable. En otras palabras, con referencia a la Figura 3c, la estrategia orientada a relación ganancia-coste se centra en el lado a la izquierda de la Figura 3c y recalcula el 18 % o el 20 % de los resultados de búsqueda precalculados con la mínima precisión más a menudo que el otro 80 % con el efecto de que el 18 % o el 20 % permanece a un nivel de precisión relativamente moderado, mientras que el otro 80 % es virtualmente insignificante. Por otra parte, el algoritmo sin tratamiento previo deja el 18 % o el 20 % de los resultados de búsqueda precalculados con la mínima precisión en un “mal estado” (es decir con muy poca precisión hasta del 0 % al 45 %), pero consigue un mejor resultado medio para el otro 80 % (Figura 3c) y a largo plazo (Figura 3a).

Por lo tanto, se busca una estrategia de recálculo que proporcione precisión superior de resultados de búsqueda precalculados a largo plazo.

Para este fin, el controlador de recálculo 2 asigna un indicador de recálculo a cualquiera de los resultados de búsqueda precalculados almacenados en la base de datos 4. Este indicador de recálculo indica la prioridad para el recálculo. El indicador de recálculo se forma de manera particular, teniendo en cuenta en general los siguientes factores:

- la probabilidad de que el resultado de búsqueda precalculado *i* sea válido o desactualizado, es decir la “precisión esperada” como se ha introducido anteriormente;
- opcionalmente, una medida de frecuencia de acceso que indica una frecuencia de solicitud desde la base de datos 4, es decir la “popularidad” como se ha introducido anteriormente;
- una frecuencia de recálculo; y
- opcionalmente una medida para los recursos de cálculo necesarios para el recálculo.

Por lo tanto, el indicador de recálculo del resultado de búsqueda recalculado *i* está basado al menos en los dos factores de la probabilidad de que el resultado de búsqueda precalculado *i* sea aún válido y de la frecuencia de recálculo del resultado de búsqueda precalculado *i*.

Más específicamente, en algunas realizaciones, el indicador de recálculo para un resultado de búsqueda precalculado *i* particular se genera ponderando la probabilidad de que el resultado de búsqueda precalculado *i* sea aún válido (“precisión esperada”) con la medida de frecuencia de acceso del resultado de búsqueda precalculado *i* que indica la frecuencia del resultado de búsqueda precalculado *i* que se solicita desde la base de datos 4 (“popularidad”) y multiplicando la frecuencia de recálculo del resultado de búsqueda *i* con la medida para los recursos de cálculo necesarios para recalcular el resultado de búsqueda precalculado *i*. En algunas realizaciones, ambos productos se dividen para formar el indicador de recálculo.

Utilizando el indicador de recálculo asociado con cualquiera de los resultados de búsqueda precalculados almacenados en la base de datos 4, el controlador de recálculo 2 selecciona estos resultados de búsqueda precalculados para recálculo que tienen un indicador de recálculo que indica la necesidad más alta para recálculo. La cantidad de resultados de búsqueda precalculados recalculados por la plataforma de cálculo recalculados en un periodo de tiempo dado, por ejemplo 20 minutos (también denominado como un “ciclo de recálculo”), está limitada por los recursos de cálculo disponibles de la plataforma de cálculo 3 para el recálculo del periodo de tiempo. Después de haber identificado los resultados de búsqueda precalculados que son más críticos de recalcular, el controlador de recálculo 2 emite una orden de recálculo a la plataforma de cálculo 3 para recalcularlos en el siguiente ciclo de recálculo. La plataforma de cálculo 3 ejecuta esta orden en una manera orientada a lotes y reenvía los resultados de búsqueda recalculados a la plataforma de búsqueda 4.

En algunas realizaciones, la plataforma de cálculo devuelve también el resultado del recálculo de vuelta al controlador de recálculo 2 al mismo tiempo. Esto posibilita que el controlador de recálculo 2 evalúe continuamente el indicador de recálculo de los resultados de búsqueda precalculados ya que están almacenados actualmente en la base de datos 4.

Un ejemplo particular de un indicador de recálculo se presenta ahora en más detalle y en una manera más formal. Este ejemplo parte de la relación ganancia-coste que se ha definido anteriormente como $a_i \frac{p_i}{p_{tot}} \frac{1 - e^{-\lambda_i t_i}}{c_i}$,

despreciando la precisión inicial a_i , como $\frac{p_i}{p_{tot}} \frac{1 - e^{-\lambda_i t_i}}{c_i}$.

- 5 Recalcular un resultado de búsqueda precalculado particular i produce un aumento en términos de aumento de precisión de usuario, en concreto la ganancia $G_i = a_i \frac{p_i}{p_{tot}} (1 - e^{-\lambda_i t_i})$.

10 Si este resultado de búsqueda precalculado i particular se recalcula más a menudo que otros resultados de búsqueda precalculados i , por ejemplo, dos veces a menudo, puede considerarse que el recálculo de i es dos veces más caro que el recálculo de i (asumiendo que un único recálculo de i y un único recálculo de i consumen la misma cantidad de recursos de cálculo de la plataforma de cálculo 3 - que no es necesariamente el caso como se explica adicionalmente a continuación). El gasto de recalculer el resultado de búsqueda precalculado i en términos del número relativo de recálculos puede definirse por lo tanto como:

15
$$\text{Gasto_de_cálculo}_i = c_i \times f_i$$

donde f_i indica una frecuencia de refresco del resultado de búsqueda precalculado i .

20 Por lo tanto, en lugar de recalculer los resultados de búsqueda precalculados con la ganancia inmediata más alta, es decir la relación ganancia-coste como se ha presentado anteriormente, se seleccionan los resultados de búsqueda precalculados con la relación ganancia-gasto más alta para recálculo, es decir con la más alta

$\frac{G_i}{\text{Gasto_de_cálculo}_i}$. Por razones de brevedad, estos resultados de búsqueda precalculados de recálculo usando este criterio de recálculo se denominan brevemente como estrategia "orientada a frecuencia de recálculo".

- 25 Una dificultad de este enfoque es que la frecuencia de recálculo f_i es en general un valor desconocido. Sin embargo, por ejemplo se aproxima por $\frac{1}{t_i}$, donde t_i indica el tiempo desde el último recálculo del resultado de búsqueda precalculado i .

Por lo tanto, $\frac{c_i}{t_i}$ puede usarse como una estimación del gasto de recálculo. La estrategia de recálculo que se ha

- 30 dirigido a un aumento a largo plazo de la precisión de los resultados de búsqueda precalculados a continuación selecciona los resultados de búsqueda precalculados con el indicador de recálculo más alto como se define como

$$\frac{G_i \times t_i}{c_i} = \frac{p_i}{p_{tot}} \frac{1 - acc_i^{t_i}}{c_i} \times t_i .$$

- 35 Los efectos de la estrategia orientada a frecuencia de recálculo presentados en el presente documento, tal como el ejemplo particular del indicador de recálculo anteriormente dado, se visualizan mediante las Figuras 4a y 4b.

40 La Figura 4a muestra la precisión de usuario conseguida por la estrategia orientada a frecuencia de recálculo (indicada por el gráfico 32) en comparación con los resultados producidos por la estrategia sin tratamiento previo de recalculer siempre los resultados de búsqueda precalculados más antiguos (indicados por el gráfico 30) y la estrategia orientada a relación coste-ganancia que incluye la consideración de la "popularidad" como se ha explicado

anteriormente (indicada por el gráfico 31), por ejemplo dado por $p_i \times \frac{1 - acc_i^{t_i}}{c_i} \times t_i$. Puede tomarse a partir de

- 45 estos gráficos 30, 31, 32 que la precisión del usuario de los resultados de búsqueda precalculados almacenados en la plataforma de búsqueda 4 se aumenta en general, a largo plazo, en aproximadamente el 7 % en comparación con la estrategia sin tratamiento previo para recalculer siempre los resultados de búsqueda precalculados más antiguos y en aproximadamente el 5 % en comparación con la estrategia orientada a relación coste-ganancia.

Por otra parte, también se mejora la precisión global por una estrategia orientada a frecuencia de recálculo que no tiene en cuenta la "popularidad" de los resultados de búsqueda precalculados, como puede inferirse a partir de la

Figura 4b. La Figura 4b está basada en la Figura 3a como se ha analizado anteriormente y visualiza adicionalmente el efecto de la estrategia orientada a frecuencia de recálculo que está basada en un indicador de recálculo de por ejemplo $\frac{1 - acc_i^{t_i}}{c_i} \times t_i$. La Figura 4b demuestra que la estrategia orientada a frecuencia de recálculo produce una

5 precisión global significativamente mejorada en comparación con la estrategia orientada a relación ganancia-coste (también empleada sin considerar la popularidad) y también una precisión global de los resultados de búsqueda precalculados anteriormente que se consiguieron por la estrategia de recálculo sin tratamiento previo de recalcular siempre los resultados de búsqueda precalculados más antiguos.

10 Opcionalmente, en algunas realizaciones, el indicador de recálculo se basa adicionalmente en el valor de precisión inicial a_i que indica la precisión esperada del resultado de búsqueda precalculado i en el momento de su recálculo, como se ha introducido adicionalmente anteriormente. En estas realizaciones, el indicador de recálculo se define,

por ejemplo, por
$$\frac{G_i \times t_i}{c_i} = \frac{p_i}{p_{tot}} \frac{a_i (1 - acc_i^{t_i})}{c_i} \times t_i$$

15 Variaciones adicionales para definir el indicador de recálculo mediante los mismos principios están dentro del alcance de la presente invención. Por ejemplo, el indicador de recálculo puede proporcionarse en una manera más genérica, con un modelo probabilístico que proporciona la probabilidad de que el resultado de búsqueda precalculado i sea inválido como $P_{invalid(i)}$, independientemente de cómo se calcule o estime $P_{invalid(i)}$. En este ejemplo más general, el indicador de recálculo se define por ejemplo como

20
$$\frac{p_i}{p_{tot}} \times P_{invalid(i)} \times \frac{t_i}{c_i}$$

Como se ha explicado anteriormente, algunas realizaciones pueden dirigirse a optimizar la precisión global sin tener en cuenta la perspectiva del usuario final, es decir la "popularidad" de los resultados de búsqueda precalculados. En estas realizaciones, el indicador de recálculo se proporciona por $P_{invalid(i)} \times \frac{t_i}{c_i}$. Como se ha descrito con

25 respecto a las realizaciones anteriores, $P_{invalid(i)}$ puede proporcionarse por $P_{invalid(i)} = (1 - e^{-\lambda_i t_i})$. En alguna realización, un factor de precisión inicial se tiene en cuenta adicionalmente de modo que $P_{invalid(i)} = a_i (1 - e^{-\lambda_i t_i})$.

30 Otras realizaciones pueden despreciar los recursos de recálculo requeridos para recalcular un resultado de búsqueda precalculado. En particular, esto se aplica a los entornos en los que cada recálculo de cualquier resultado de búsqueda precalculado requiere la misma cantidad de recursos de cálculo.

35 Algunas realizaciones presentan un indicador de recálculo perfeccionado adicional dirigido a una estrategia de recálculo para optimización a largo plazo de la precisión de usuario que implementa la estrategia orientada a frecuencia de recálculo como se ha deducido anteriormente. Este indicador de recálculo perfeccionado corresponde a la siguiente expresión:

$$-\frac{p_i}{c_i} \left(t_i e^{-\lambda_i t_i} + \frac{e^{-\lambda_i t_i} - 1}{\lambda_i} \right)$$

40 Esto se deriva matemáticamente como sigue:

Un conjunto de frecuencias de recálculo se define $\{f_i\}_{i \in dd}$ de un resultado de búsqueda precalculado i . Se definen periodos correspondientes entre recálculo del resultado de búsqueda precalculado i como $T_i = \frac{1}{f_i}$.

45 Si se aplica la frecuencia f_j al resultado de búsqueda precalculado j , entonces la precisión media de j se proporciona mediante

$$\overline{Acc_i} = \frac{1}{T_i} \int_0^{T_i} e^{-\lambda_i t} dt = \frac{(1 - e^{-\lambda_i T_i})}{\lambda_i T_i}$$

La precisión de usuario media se define entonces como

$$\overline{\text{Precisión_de_usuario}} = \sum_{x_i \in dd} \frac{p_i}{p_{tot}} \overline{Acc_i} = \sum_{x_i \in dd} \frac{p_i}{p_{tot}} \frac{(1 - e^{-\lambda_i T_i})}{\lambda_i T_i}$$

Si los recursos de cálculo disponibles para un periodo de tiempo son iguales para cualquier periodo de tiempo entonces:

$\sum_{x_i \in dd} c_i f_i = R$ donde R son los recursos de cálculo asignados por periodo de tiempo y x_i indica los resultados de búsqueda precalculados. Sustituir la frecuencia de recálculo f_i con los periodos de tiempo T_i entre los recálculos, se deduce que

$$\sum_{x_i \in dd} c_i f_i = \sum_{x_i \in dd} \frac{c_i}{T_i} = R$$

El conjunto de $\{(T_i)_{x_i \in dd}, \sum_{x_i \in dd} \frac{c_i}{T_i} = R\}$ puede indicarse como H .

Añadir un multiplicador de Lagrange da como resultado una definición de

$$l(\theta, (T_i)_{x_i \in dd}) = \sum_{x_i \in dd} \frac{p_i}{p_{tot}} \frac{(1 - e^{-\lambda_i T_i})}{\lambda_i T_i} + \theta \left(R - \sum_{x_i \in dd} \frac{c_i}{T_i} \right)$$

En el conjunto de H , se aplica que $l(\theta, (T_i)_{x_i \in dd}) = \overline{\text{Precisión_de_usuario}}$.

Por lo tanto, la precisión de usuario e l tienen el mismo máximo en H . Y sucede que el máximo de $l(\theta, (T_i)_{x_i \in dd})$ se da en H puesto l tiene su máximo para

$$\frac{\partial l}{\partial \theta} = R - \sum_{x_i \in dd} \frac{c_i}{T_i} = 0.$$

Por lo tanto, hallando el máximo de $\overline{\text{Precisión_de_usuario}}$ en H es equivalente a hallar el máximo de l

$$\begin{aligned} \frac{\partial l}{\partial T_i} &= \frac{\partial}{\partial T_i} \left(\frac{p_i}{p_{tot}} \frac{1 - e^{-\lambda_i T_i}}{\lambda_i T_i} \right) - \theta \frac{\partial}{\partial T_i} \left(\frac{c_i}{T_i} \right) \\ \frac{\partial l}{\partial T_i} &= -\frac{p_i}{p_{tot}} \frac{e^{-\lambda_i T_i}}{T_i} + \left(\frac{p_i}{p_{tot}} \frac{1 - e^{-\lambda_i T_i}}{\lambda_i T_i^2} \right) + \theta \frac{c_i}{T_i^2} \end{aligned}$$

El máximo se halla si $\frac{\partial l}{\partial T_i} = 0$.

Si multiplicamos por $\frac{p_{tot} T_i^2}{c_i}$: tenemos un máximo para

$$-\frac{p_i}{c_i} \left(T_i e^{-\lambda_i T_i} + \frac{1 - e^{-\lambda_i T_i}}{\lambda_i} \right) + \theta p_{tot} = 0$$

es decir

$$\mathcal{F}_i(T_i) \triangleq -\frac{p_i}{c_i} \left(T_i e^{-\lambda_i T_i} + \frac{e^{-\lambda_i T_i} - 1}{\lambda_i} \right) = \text{const}$$

5 La estrategia de recálculo que optimiza el indicador de precisión de usuario puede a continuación emplearse como sigue: los resultados de búsqueda precalculados con la $\mathcal{F}_i(t)$ más alta como se define en este punto se han de recalcular en primer lugar para tener la $\mathcal{F}_i(t)$ tan igual como sea posible (\mathcal{F}_i es una función creciente).

Es interesante comparar este indicador de recálculo perfeccionado con el indicador de recálculo anterior descrito para la estrategia orientada a frecuencia de recálculo anterior, por ejemplo $p_i \frac{1 - acc_i'}{c_i} \times t_i$ que se ha introducido específicamente anteriormente:

$$\mathcal{F}_i(T_i) - \frac{p_i (1 - e^{-\lambda_i T_i})}{c_i} t_i = \frac{p_i (1 - \lambda_i T_i - e^{-\lambda_i T_i})}{c_i \lambda_i} = O\left(\frac{p_i}{c_i} \lambda_i T_i^2\right)$$

15 Esto significa que para frecuencias de recálculo no demasiado bajas que son de una cantidad significativa (es decir el periodo de tiempo desde el último recálculo no es demasiado largo), el indicador de recálculo introducido adicionalmente anteriormente es una buena aproximación del indicador de recálculo presentado adicionalmente en este punto.

20 Adicionalmente, si se vuelve a introducir la precisión inicial, el indicador de recálculo perfeccionado se proporciona por

$$-\frac{a_i p_i}{c_i} \left(t_i e^{-\lambda_i t_i} + \frac{e^{-\lambda_i t_i} - 1}{\lambda_i} \right).$$

25 La estrategia orientada a frecuencia de recálculo se ha analizado hasta ahora asumiendo que el recálculo para cualquier resultado de búsqueda precalculado por la plataforma de cálculo 3 requiere sustancialmente la misma cantidad de recursos de cálculo independientemente de si los resultados de búsqueda precalculados se calculan por separado, junto con resultados de búsqueda adyacentes (por ejemplo mismo origen, destino y fechas adyacentes) o cualquier otro resultado de búsqueda no adyacente. En la práctica, sin embargo, esta suposición no puede hacerse en general puesto que, por ejemplo, ciertos resultados de búsqueda precalculados y/o sus datos de cálculo subyacentes correspondientes están interrelacionados entre sí. Recalcular tales resultados de búsqueda precalculados interrelacionados juntos (es decir, en el mismo ciclo de recálculo) podría incluir efectos sinérgicos y puede ser por lo tanto más eficaz que recalcularlos por separado. Por lo tanto, cuando se recalculan resultados de búsqueda precalculados interrelacionados junto con los costes de cálculo por resultado de búsqueda precalculado individual es en general menor que el coste de cálculo por resultado de búsqueda precalculado individual requerido si se recalculan por separado, es decir en diferentes ciclos de recálculo. La estrategia orientada a frecuencia de recálculo anterior se ha deducido sin tener en cuenta tales efectos sinérgicos de recalcular ciertos resultados de búsqueda precalculados interrelacionados como un paquete total en un único ciclo de recálculo.

40 Para fomentar el entendimiento de los resultados de búsqueda precalculados interrelacionados, se proporciona a continuación un ejemplo particular. Este ejemplo se origina de una aplicación particular del entorno de base de datos distribuido 1 en la industria de los viajes, es decir la base de datos 4 mantiene resultados de búsqueda precalculados relacionados con viajes y los hace disponibles para usuarios finales. El siguiente ejemplo, sin embargo, no se asume para limitar el problema de resultados de búsqueda precalculados interrelacionados tales como una aplicación de datos de viaje. En su lugar, están presentes condiciones similares o análogas que permiten un recálculo sinérgico y por lo tanto más eficaz de resultados de búsqueda precalculados interrelacionados en sistemas de base de datos independientes del contenido de conjuntos de datos precalculados. En general, puede decirse que cualquier proceso de recálculo de resultados de búsqueda precalculados tendrá por objetivo una mutualización de subtareas de recálculo que se han de ejecutar comúnmente para cualquier resultado de búsqueda precalculado de un conjunto de resultados de búsqueda precalculados. Por lo tanto, recalcular resultados de búsqueda precalculados juntos que tienen tal tarea de recálculo en común, es en general favorable sobre recalcular solicitudes de búsqueda precalculadas juntas que no comparten subtareas de recálculo similares.

55 En una realización particular, las solicitudes de búsqueda precalculadas son registros de datos de vuelo de ida y vuelta, especificando cada uno un origen y destino de viaje y una fecha de salida y llegada (o, como alternativa a la fecha de llegada, una duración de permanencia relacionada con la fecha de salida). En este ejemplo, la base de

5 datos 4 contiene recomendaciones de viaje de ida y vuelta precalculadas para que se cubra cualquier par origen-destino y cualquier par fecha-salida-llegada. La Tabla 1 indica un pequeño fragmento a partir de las recomendaciones de viaje precalculadas mantenidas en la base de datos 4, siendo el fragmento recomendaciones de viaje para el par de ciudades Niza-Boston (“NCE-BOS”) y para fechas de salida del 1 de julio al 5 de julio con duración de permanencia máxima de cinco días, la abreviatura “pc-fr x” significa “número de recomendación de viaje precalculado x”.

Tabla 1

NCE-BOS	Salida 1 JUL	Salida 2 JUL	Salida 3 JUL	Salida 4 JUL	Salida 5 JUL
Recuperado 2 JUL	pc-fr 0	--	--	--	--
Recuperado 3 JUL	pc-fr 1	pc-fr 5	--	--	--
Recuperado 4 JUL	pc-fr 2	pc-fr 6	pc-fr 10	--	--
Recuperado 5 JUL	pc-fr 3	pc-fr 7	pc-fr 11	pc-fr 15	--
Recuperado 6 JUL	pc-fr 4	pc-fr 8	pc-fr 12	pc-fr 16	pc-fr 20
Recuperado 7 JUL	--	pc-fr 9	pc-fr 13	pc-fr 17	pc-fr 21
Recuperado 8 JUL	--	--	pc-fr 14	pc-fr 18	pc-fr 22
Recuperado 9 JUL	--	--	--	pc-fr 19	pc-fr 23
Recuperado 10 JUL	--	--	--	--	pc-fr 24

10 El recálculo de una recomendación de viaje precalculada particular, por ejemplo pc-fr 9 con relación a una fecha de salida del 2 de julio y una fecha de retorno del 7 de julio, se realiza de manera aproximada por la plataforma de cálculo 3 mediante las siguientes subtareas:

15 1: recuperar las posibles rutas geográficas entre Niza y Boston que incluyen paradas/intercambios intermedios, por ejemplo: Niza-París-Boston, Niza-Londres-Boston, Niza-Nueva York-Boston, Niza-París-Nueva York-Boston, etc.

20 2: hallar todos los posibles vueltos en todas las rutas recuperadas por la actividad 1: por ejemplo Air France 123 y Emirates 7123 en el tramo Niza-Paris, Air France 456 y American Airlines 564 y United Airlines 125 en el tramo París-Nueva York, etc.

25 a. esta actividad se realiza para la parte de ida del viaje en la fecha de salida del 2 de julio (subtarea 2a) y
b. para la parte de retorno del viaje en la fecha de retorno del 7 de julio (subtarea 2b) puesto que los vuelos generalmente no están disponibles todos los días.

30 3: comprobar la disponibilidad en todos los vuelos determinados por las actividades 2a y 2b, es decir el 2 de julio para la parte de salida del recorrido (= subtarea 3a) y el 7 de julio para la parte de retorno del recorrido (= subtarea 3b).

35 4: recuperar tarifas que pueden aplicarse a estos vuelos en estas fechas. Una tarifa típica es la regla que produce un precio para todo el recorrido. Sin embargo, las tarifas pueden ser también válidas para la parte de salida y para la de entrada del recorrido en sus fechas respectivas caso en el que pueden recuperarse una vez para la parte de salida (= subtarea 4a) y una vez para la parte de retorno del recorrido (= subtarea 4b). Las tarifas tienen restricciones en las fechas de salida, en las fechas de retorno, en los en los que se aplican y muchas otras. Las tarifas pueden combinarse juntas, descontarse en algunos casos específicos y así sucesivamente.

5: la mejor solución, es decir la relación con el mejor precio, se mantiene, se devuelve por la plataforma de cálculo 3 y se almacena por la base de datos 4 como pc-fr 9.

40 Estas subtareas se realizan de una forma similar para cualquier recomendación de viaje precalculada. Cuando se realiza el recálculo para todo el conjunto de recomendaciones de viaje precalculadas (pc-fr 0 a pc-fr 24), pueden mutualizarse varias subtareas de recálculo. Ejecutar la subtarea 1 es necesario únicamente una vez para todo el conjunto de recomendaciones de viaje precalculadas pc-fr 0 a pc-fr 24 puesto que todas estas recomendaciones de viaje precalculadas pertenecen al mismo par de ciudades origen-destino Niza-Boston. Por lo tanto, la subtarea 1 no
45 necesita rehacer para cada resultado de búsqueda precalculado pc-fr 0 a pc-fr 24.

Las subtareas 2a, 3a y la subtarea 4 son, por otra parte, específicas a una fecha de salida. Pueden por lo tanto reutilizarse para todas las recomendaciones de viaje precalculadas relacionadas con una y la misma fecha de salida. La Tabla 2 indica esto para las recomendaciones de viaje precalculadas pc-fr 5 a pc-fr 9 relacionadas con la fecha

de salida del 2 de julio:

Tabla 2

NCE-BOS	Salida 1 JUL	Salida 2 JUL	Salida 3 JUL	Salida 4 JUL	Salida 5 JUL
Recuperado 2 JUL	pc-fr 0	--	--	--	--
Recuperado 3 JUL	pc-fr 1	Subtareas comunes: 2a, 3a, 4a	--	--	--
Recuperado 4 JUL	pc-fr 2		pc-fr 10	--	--
Recuperado 5 JUL	pc-fr 3		pc-fr 11	pc-fr 15	--
Recuperado 6 JUL	pc-fr 4		pc-fr 12	pc-fr 16	pc-fr 20
Recuperado 7 JUL	--		pc-fr 13	pc-fr 17	pc-fr 21
Recuperado 8 JUL	--	--	pc-fr 14	pc-fr 18	pc-fr 22
Recuperado 9 JUL	--	--	--	pc-fr 19	pc-fr 23
Recuperado 10 JUL	--	--	--	--	pc-fr 24

5 Análogamente, las subtareas 2b, 3b y 4b son específicas a una fecha de retorno, y por lo tanto, se realizan comúnmente para recomendaciones de viaje precalculadas relacionadas con una y la misma fecha de retorno. Esto se ilustra mediante la tabla 3 para las recomendaciones de viaje precalculadas pc-fr 9, pc-fr 13, pc-fr 17 y pc-fr 21, todas las cuales hacen referencia a la fecha de retorno del 7 de julio:

10

Tabla 3

NCE-BOS	Salida 1 JUL	Salida 2 JUL	Salida 3 JUL	Salida 4 JUL	Salida 5 JUL
Recuperado 2 JUL	pc-fr 0	--	--	--	--
Recuperado 3 JUL	pc-fr 1	pc-fr 5	--	--	--
Recuperado 4 JUL	pc-fr 2	pc-fr 6	pc-fr 10	--	--
Recuperado 5 JUL	pc-fr 3	pc-fr 7	pc-fr 11	pc-fr 15	--
Recuperado 6 JUL	pc-fr 4	pc-fr 8	pc-fr 12	pc-fr 16	pc-fr 20
Recuperado 7 JUL	--	Subtareas comunes: 2b, 3b, 4b			
Recuperado 8 JUL	--	--	pc-fr 14	pc-fr 18	pc-fr 22
Recuperado 9 JUL	--	--	--	pc-fr 19	pc-fr 23
Recuperado 10 JUL	--	--	--	--	pc-fr 24

15 Por lo tanto, en resumen, únicamente una parte de la subtarea 4, en concreto recuperar tales tarifas que no son válidas para la parte de salida total del viaje y para la parte de retorno total del viaje, pero son específicas para subconjuntos o recomendaciones de viaje particulares, tienen que realizarse por separado para cada recomendación de viaje precalculada, aunque pueden realizarse otras subtareas en común para todas las recomendaciones de viaje precalculadas relacionadas con el mismo par de ciudades origen-destino (se cumple para la subtarea 1) o al menos para las recomendaciones de viaje precalculadas relacionadas con la misma fecha de salida (subtareas 2a, 3a y 4a) o a la misma fecha de retorno (subtareas 2b, 3b y 4b). En consecuencia, cuantas más recomendaciones de viaje precalculadas estén relacionadas con un par de ciudades origen-destino y cuantas más recomendaciones de viaje precalculadas estén relacionadas a una fecha de salida y fecha de retorno, respectivamente, más recursos de cálculo pueden liberarse mutualizando estas subtareas a través de las respectivas solicitudes de vuelo precalculadas.

25 Esta relación entre aumentar los efectos sinérgicos y aumentar el número de recomendaciones de viaje precalculadas con relación a un par de ciudades origen-destino y a una fecha de salida o de retorno se visualiza también mediante la Figura 5. La Figura 5 muestra seis gráficos de conjuntos de recomendación de viaje precalculados ejemplares, perteneciendo cada conjunto a un par de ciudades origen-destino. Estando asociadas en número decreciente de recomendaciones de viaje precalculadas con un par de ciudades, el gráfico 36 se refiere al par de ciudades Nueva York-Búfalo, el gráfico 37 a Nueva York-Minsk, el gráfico 38 a Nueva York-Hilo en Hawái, el gráfico 39 a Nueva York-Bilbao, el gráfico 40 a Nueva York-Male y, finalmente, el gráfico 41 a Nueva York-Mauricio. El eje X del diagrama de la Figura 5 indica el número de recomendaciones de viaje precalculadas, mientras que el eje Y representa una medida de recursos de recálculo necesarios para realizar un recálculo de recomendaciones de

30

viaje precalculadas, en concreto tiempo de CPU.

Por otra parte, la Figura 5 muestra que el recálculo de algunos conjuntos de recomendaciones de viaje precalculadas requieren más recursos de cálculo que otros. Por ejemplo, recalculando recomendaciones de viaje precalculadas a partir del conjunto 41 que incluye recomendaciones de viaje precalculadas para vuelos entre Nueva York y Mauricio es en general más costoso que recalculando recomendaciones de viaje precalculadas a partir del conjunto 36 que incluye recomendaciones de viaje precalculadas para vuelos entre Nueva York y Búfalo. Esta diferencia general de la cantidad de recursos de cálculo necesaria para recalculando recomendaciones de viaje precalculadas se tiene en cuenta incluyendo los costes de recálculo C en el indicador de recálculo como ya se ha explicado en detalle anteriormente. Sin embargo, los gráficos de la Figura 5 indican también que los costes de recálculo varían con el número de recomendaciones de viaje precalculadas de un conjunto que se está recalculando. En general, cuantas más recomendaciones de viaje precalculadas relacionadas con un conjunto se recalculan juntas, menos recursos de cálculo son necesarios para recalculando cada recomendación de viaje precalculada. Por ejemplo, el gráfico 36 indica que recalculando 32 recomendaciones de viaje precalculadas del conjunto Nueva York-Búfalo requiere aproximadamente 1000 ms de CPU, es decir aproximadamente 31,25 ms de CPU por recomendación de viaje precalculada, mientras que recalculando 170 recomendaciones de viaje precalculadas requiere aproximadamente 5000 ms de CPU, es decir únicamente aproximadamente 29,4 ms de CPU por recomendación de viaje precalculada. Con respecto al gráfico 38 relacionado con el par de ciudades Nueva York-Hilo, recalculando 35 recomendaciones de viaje precalculadas a partir de ese conjunto requiere aproximadamente 10.000 ms de CPU, es decir 286 ms de CPU por recomendación de viaje precalculada, mientras que recalculando 140 recomendaciones de viaje precalculadas a partir de ese conjunto requiere aproximadamente 25.000 ms de CPU, es decir únicamente 178,5 ms de CPU por recomendación de viaje precalculada. Por lo tanto, es evidente que las recomendaciones de viaje precalculadas que pertenecen a un par de ciudades origen-destino están interrelacionadas en términos de recursos informáticos requeridos para su recálculo como se ha explicado anteriormente.

En otras palabras, los recursos de cálculo necesarios para recalculando un resultado de búsqueda precalculado i generalmente dependen de si se recalculan o no otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i durante el mismo ciclo de cálculo. Por lo tanto, los recursos de cálculo para recalculando los resultados de búsqueda precalculados no son estáticos, sino que varían con la selección del conjunto de resultados de búsqueda precalculados para que se recalculen durante el ciclo de cálculo.

Estos recursos de cálculo que varían con el número de resultados de búsqueda precalculados interrelacionados que se recalculan juntos se tienen en cuenta por la estrategia de recálculo empleada por algunas realizaciones como sigue:

En general, los recursos de cálculo necesarios para recalculando los resultados de búsqueda precalculados a recalculando se estiman dinámicamente por el controlador de recálculo 2 mientras se seleccionan los resultados de búsqueda precalculados a recalculando durante el siguiente ciclo de cálculo. Esta estimación depende de qué otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i se seleccionen para el recálculo durante el siguiente ciclo de recálculo. En algunas realizaciones, esto se consigue mediante una estimación perfeccionada de manera iterativa de los recursos de cálculo necesarios para recalculando los resultados de búsqueda precalculados a recalculando mientras se determina el subconjunto de los resultados de búsqueda precalculados a recalculando realmente. Esta estimación iterativa de los recursos de cálculo variables incluye las siguientes actividades:

a) para cualquier resultado de búsqueda precalculado i, los recursos de cálculo c_i necesarios para recalculando el resultado de búsqueda precalculado i se inicializan con un primer valor aproximado. Este valor asume que el recálculo del resultado de búsqueda precalculado i es independiente del cálculo de otros resultados de búsqueda precalculados seleccionados para recálculo durante el siguiente ciclo de recálculo.

b) una porción de los resultados de búsqueda precalculados para recálculo se selecciona a continuación. Esta selección se hace, por ejemplo, de acuerdo con el indicador de recálculo como ya se ha explicado anteriormente. La porción seleccionada, sin embargo, no ha agotado ya los recursos de cálculo disponibles completos de la plataforma de cálculo 3 al 100 %, sino que únicamente consume una parte de los recursos de cálculo disponibles para el siguiente ciclo de recálculo. Por ejemplo, en algunas realizaciones, la porción seleccionada únicamente requiere un porcentaje dado de los recursos de cálculo disponibles de la plataforma de cálculo 3 para recálculo en el siguiente ciclo de recálculo. En las realizaciones, se usan porcentajes específicos como el porcentaje dado, tales como 1 %, 2 %, 5 %, 10 %, 20 %, 25 %, 30 %, 40 %, 50 % o 60 % o valores más altos por debajo del 100 %. En cualquier caso, la selección de esta porción de resultados de búsqueda precalculados a recalculando en el siguiente ciclo de recálculo es basándose en los valores actuales para los recursos de cálculo necesarios para recalculando la porción de resultados de búsqueda precalculados, es decir en la primera iteración de selección basándose en los valores de la actividad de inicialización a), es decir sin tener en cuenta ninguna interrelación o dependencia entre los resultados de búsqueda precalculados a recalculando.

c) para cualquier resultado de búsqueda precalculado i (incluyendo tanto los resultados de búsqueda precalculados seleccionados por la actividad b) anterior como los resultados de búsqueda precalculados que no se han seleccionado por la actividad b)), el controlador de recálculo 2 vuelve a evaluar los recursos de cálculo c_i necesarios para recalcular el resultado de búsqueda precalculado i teniendo en cuenta qué resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i se han seleccionado para recálculo en la actividad c). Para los resultados de búsqueda precalculados que ya se han seleccionado por la actividad b), esta reevaluación proporciona valores perfeccionados de c_i y en total un valor perfeccionado del porcentaje de los recursos de cálculo disponibles de la plataforma de cálculo para recálculo en el siguiente ciclo de recálculo necesario para recalcular los resultados de búsqueda precalculados seleccionados para recálculo en el siguiente ciclo de cálculo hasta ahora. Hasta el punto que los resultados de búsqueda precalculados interrelacionados se hayan seleccionado en la iteración actual, el valor perfeccionado de $\sum c_i$ de los resultados de búsqueda precalculados ya seleccionados es generalmente menor que el valor (teórico) de $\sum c_i$ de los resultados de búsqueda precalculados ya seleccionados despreciando su interrelación y asumiendo un recálculo del precálculo ya seleccionado sin recalcular ningún resultado de búsqueda precalculado interrelacionado. La reevaluación de los resultados de búsqueda precalculados que no se han seleccionado (aún) para recálculo es detectable puesto que son todos candidatos para una selección en la siguiente iteración o iteraciones. Si uno o más resultados de búsqueda precalculados interrelacionados a resultados de búsqueda precalculados no seleccionados aún ha/han sido seleccionados en la actividad anterior b), los recursos de cálculo necesarios para recalcular estos resultados de búsqueda precalculados no seleccionados aún con resultado o resultados de búsqueda precalculado interrelacionados son en general inferiores (y por lo tanto en general se han de reducir por esta actividad c)) que los recursos de cálculo requeridos si no se seleccionó resultado de búsqueda precalculado interrelacionado para el precálculo.

d) el controlador de recálculo 2 hace referencia de vuelta a la actividad b) si se agota menos del 100 % de los recursos de cálculo totales disponibles de la plataforma de cálculo para recálculo en el siguiente ciclo de recálculo.

En general, este enfoque es independiente de la manera específica de cómo se calcula el indicador de recálculo o de qué clase de estrategia de recálculo se emplee.

La Figura 6 muestra un ejemplo relacionado con viajes de un entorno de base de datos 1 que representa detalles adicionales en comparación con la Figura 1. El controlador de recálculo 2 mantiene una representación de los resultados de búsqueda precalculados almacenados por la plataforma de búsqueda 4, por ejemplo, en la base de datos 6. El controlador de recálculo 2 controla el recálculo de los resultados de búsqueda precalculados empleando el indicador de recálculo como se ha descrito en detalle anteriormente. Los resultados de búsqueda precalculados a recalcular se ordenan para recalcularse por órdenes de cálculo que el controlador de recálculo 2 transmite a la plataforma de cálculo 3. La plataforma de cálculo recalcula los respectivos resultados de búsqueda precalculados. Como se representa mediante el entorno ejemplar 1 de la Figura 6, el recálculo realizado por la re-plataforma de cálculo 3 puede ser basándose en datos subyacentes tales como tarifas mantenidas en una base de datos de tarifas 7, horarios de transporte mantenidos en la base de datos de horarios 8 y datos de disponibilidad de transporte mantenidos en la base de datos de disponibilidad 9. La plataforma de recálculo 3 envía los resultados de búsqueda recalculados a la plataforma de búsqueda 4 y los devuelve al controlador de recálculo 2. Como se ha señalado anteriormente, el controlador de recálculo 2 puede estar integrado con la plataforma de cálculo 3 y/o la plataforma de búsqueda 4. Los resultados de búsqueda precalculados actualizados de esta manera se solicitan por una aplicación de búsqueda 5 desde la plataforma de búsqueda 6 por ejemplo usando interfaces de servicio web.

Ahora haciendo referencia a la Figura 7, algunas realizaciones emplean una estructura modular del controlador de recálculo 2 para conseguir los métodos anteriormente descritos. Algunas de estas partes ya se han descrito en la solicitud internacional no publicada PCT/EP2013/002390 a la que se hace referencia para una explicación más detallada de la estructura del controlador de recálculo. Algunos detalles más con respecto a la lógica interna del controlador de recálculo y sus relaciones se elaboran en este punto. Como se muestra mediante la Figura 7, el controlador de recálculo 2 incluye de manera ejemplar los siguientes componentes:

- el componente de representación de datos interno 10: este componente proporciona herramientas para crear, almacenar, actualizar y acceder a grandes matrices que representan los resultados de búsqueda precalculados almacenados en la base de datos 4. La función principal del componente de representación de datos interno 10 es proporcionar un "espejo" de los resultados de búsqueda precalculados almacenados en la base de datos 4 que sirve como la base para analizar los resultados de búsqueda precalculados para decidir cuáles de ellos se han de recalcular en el siguiente ciclo de recálculo. Con mayor precisión, el componente de representación de datos interno 10 no mantiene una copia de uno a uno de los resultados de búsqueda precalculados como se almacenan en la base de datos 4, sino una representación apropiada que no tiene que incluir todos los detalles de los resultados de búsqueda precalculados como se almacenan en la base de datos 4, pero, por otra parte, incluye datos de control adicionales asociados con los resultados de búsqueda precalculados tal como los tiempos de su último recálculo y, en particular, el indicador de recálculo.

- gestor de entrada 11: este componente introduce datos desde orígenes heterogéneos tales como una base de datos u origen de datos de tasa de validez, una base de datos u origen de datos de popularidad, una base de datos u origen de datos de precisión inicial, una base de datos u origen de datos de costes y/u orígenes que indican eventos en tiempo real que potencialmente influyen la validez de los resultados de búsqueda precalculados. Estos datos se usan, por ejemplo, para generar y actualizar los indicadores de recálculos asociados con los resultados de búsqueda precalculados como se ha explicado en detalle anteriormente. El gestor de entrada 11 convierte los datos entrantes en los formatos de datos apropiados y actualiza matrices correspondientes que representen los resultados de búsqueda precalculados como se almacenan por el componente de representación de datos interno 10.
- Analizador 12: este componente calcula matrices de datos intermedias implicadas por el modelo probabilístico (precisión, criticidad) basándose en las matrices almacenadas por el componente de representación de datos interno 10.
- Gestor de eventos 13: este componente agrega información sobre información de eventos en tiempo real y modifica las predicciones de validez proporcionadas por el modelo probabilístico en consecuencia.
- Optimizador 14: este componente ejecuta la estrategia de recálculo, es decir el recálculo orientado a frecuencia de recálculo y la selección iterativa de resultados de búsqueda precalculados teniendo en cuenta costes de cálculo variables de resultados de búsqueda precalculados interrelacionados como se ha descrito en detalle anteriormente. Después de haber determinado los resultados de búsqueda precalculados a recalcular, el optimizador 14 genera órdenes de recálculo y las emite a la plataforma de cálculo 3. Adicionalmente, actualiza el tiempo de recálculo de estos resultados de búsqueda precalculados almacenados en el componente de representación de datos interno 10.

Los últimos dos módulos, el gestor de eventos 13 y el optimizador 14, se agrupan bajo el nombre "consolidador" en el documento PCT/EP2013/002390.

Finalmente, la Figura 8 es una representación esquemática de un sistema informático que proporciona la funcionalidad del controlador de recálculo 2 como se muestra por las Figuras 2, 6 y 7. En el controlador de recálculo 2 pueden ejecutarse un conjunto de instrucciones, para provocar que el sistema informático realice cualquiera de los métodos analizados en el presente documento. El controlador de recálculo 2 incluye un procesador 101, una memoria principal 102 y un dispositivo de interfaz de red 103, que comunican entre sí mediante un bus 104. Opcionalmente, puede incluir adicionalmente una memoria estática 105 y una unidad de disco 106. Una pantalla de vídeo 107, un dispositivo de entrada alfanumérico 108 y un dispositivo de control de cursor 109 pueden formar una interfaz de usuario de navegador de lista de distribución. El dispositivo de interfaz de red 103 conecta el controlador de recálculo de datos 2 a la plataforma de cálculo 3, los orígenes de datos estadísticos necesarios para rellenar el modelo predictivo tales como servidores de estadísticas, una base de datos u origen de datos de volatilidad y una base de datos u origen de datos de precisión inicial, los orígenes de eventos en tiempo real, internet y/o cualquier otra red. Un conjunto de instrucciones (es decir software) 110 que realizan uno o todos los métodos anteriormente descritos, residen completa, o al menos parcialmente, dentro o en un medio legible por máquina, por ejemplo la memoria principal 102 y/o el procesador 101. Un medio legible por máquina en el que reside el software 110 puede también ser un soporte de datos no volátil 111 (por ejemplo un disco duro magnético no extraíble o un disco extraíble óptico o magnético) que es parte de la unidad de disco 106. El software 110 puede transmitirse o recibirse adicionalmente como una señal propagada 112 mediante internet a través del dispositivo de interfaz de red 103.

La presente estrategia de recálculo proporciona un medio para generar automáticamente decisiones de recálculo que se dirigen a mejorar la validez de resultados de búsqueda precalculados. Determina qué resultados de búsqueda precalculados se han de recalcular y controla el recálculo también por tiempo teniendo en cuenta los recursos de cálculo disponibles en la plataforma de cálculo. Por lo tanto, en general, la precisión/validez de los resultados de búsqueda se estima en el modelo probabilístico que modela la actualización y desactualización, respectivamente, durante el tiempo y tiene en cuenta una frecuencia de recálculo de los resultados de búsqueda precalculados. Los resultados de búsqueda precalculados que se han de recalcular más a menudo que otros se considera que son más "caros" de mantener actualizados. De esta manera, se evita un enfoque ineficaz en recalcular constantemente resultados de búsqueda precalculados muy volátiles.

REIVINDICACIONES

1. Un método para determinar resultados de búsqueda precalculados a recalcularse, realizándose el método en un entorno de base de datos (1), comprendiendo el entorno de base de datos una plataforma de búsqueda (4) que mantiene resultados de búsqueda precalculados, un controlador de recálculo (2) y una plataforma de cálculo (3), estando el método **caracterizado por** comprender:

- el controlador de recálculo (2) seleccionar de manera iterativa un número de los resultados de búsqueda precalculados a recalcularse por la plataforma de cálculo (3) durante un intervalo de tiempo dado que forma un ciclo de recálculo, el número de resultados de búsqueda precalculados estando limitado por los recursos de cálculo de la plataforma de cálculo disponibles durante el intervalo de tiempo dado, en el que

o los recursos de cálculo c_i necesarios para recalcularse un resultado de búsqueda precalculado i dependen de si se recalculan o no otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i durante el intervalo de tiempo dado y en el que

o los recursos de cálculo c_i necesarios para recalcularse el resultado de búsqueda precalculado i se estiman dinámicamente y de manera iterativa dependiendo de qué otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i se seleccionan para recálculo durante el intervalo de tiempo dado;

en el que seleccionar de manera iterativa un número de los resultados de búsqueda precalculados a recalcularse por la plataforma de cálculo (3) durante un intervalo de tiempo dado y estimar dinámicamente y de manera iterativa los recursos de cálculo c_i necesarios para recalcularse resultado de búsqueda precalculado i durante el intervalo de tiempo dado comprende:

a) inicializar los recursos de cálculo c_i necesarios para recalcularse cada resultado de búsqueda precalculado i con un valor que supone el cálculo de i independiente del cálculo de otros resultados de búsqueda precalculados seleccionados para recálculo durante el intervalo de tiempo dado;

b) en una primera iteración de selección, seleccionar una primera porción de los resultados de búsqueda precalculados para recálculo, en el que - basándose en los valores actuales para los recursos de cálculo necesarios para recalcularse la porción de resultados de búsqueda precalculados - la primera porción seleccionada únicamente requiere un porcentaje por debajo del 100 % de los recursos de cálculo globales de la plataforma de cálculo disponibles para recálculo en el intervalo de tiempo dado;

c) reevaluar los recursos de cálculo c_i necesarios para recalcularse cada resultado de búsqueda precalculado i teniendo en cuenta qué resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i se han seleccionado para recálculo;

d) realizar iteraciones adicionales de selección de porciones de resultados de búsqueda precalculados aún no seleccionados adicionales y reevaluar los recursos de cálculo c_i necesarios para recalcularse cada resultado de búsqueda precalculado i si se agota menos del 100 % de los recursos de cálculo globales de la plataforma de cálculo (3) disponibles para recálculo en el intervalo de tiempo dado;

- el controlador de recálculo (2) generar y transmitir órdenes de recálculo para recalcularse el número de resultados de búsqueda precalculados a la plataforma de cálculo (3).

2. El método de la reivindicación 1, en el que reevaluar los recursos de cálculo c_i necesarios para recalcularse cada resultado de búsqueda precalculado i incluye proporcionar un valor perfeccionado del porcentaje de los recursos de cálculo globales de la plataforma de cálculo disponibles para recálculo en el intervalo de tiempo dado necesarios para recalcularse la porción seleccionada de los resultados de búsqueda precalculados en el intervalo de tiempo dado.

3. El método de la reivindicación 1 o la reivindicación 2, en el que seleccionar el número de resultados de búsqueda precalculados para recálculo comprende adicionalmente:

el controlador de recálculo (2) asignar un indicador de recálculo a cualquiera de los resultados de búsqueda precalculados, en el que el indicador de recálculo para el resultado de búsqueda precalculado i indica una necesidad para recálculo del resultado de búsqueda precalculado i durante el intervalo de tiempo dado y está basado al menos en los siguientes factores:

o una probabilidad acc_i de que el resultado de búsqueda precalculado i sea aún válido;

o una frecuencia de recálculo f_i del resultado de búsqueda precalculado i aproximada por la recíproca de un tiempo desde el último recálculo del resultado de búsqueda precalculado i ; y

o una medida para los recursos de cálculo estimados c_i necesarios para recalcularse el resultado de búsqueda precalculado i ;

- seleccionar la primera porción y las porciones de los resultados de búsqueda precalculados aún no seleccionados para recálculo de acuerdo con el indicador de recálculo.

4. El método de la reivindicación 3, en el que el indicador de recálculo se define por $\frac{1 - acc_i}{f_i c_i}$.

5. El método de la reivindicación 3 o la reivindicación 4, en el que el indicador de recálculo está basado adicionalmente en una frecuencia de acceso p_i del resultado de búsqueda precalculado i desde la plataforma de

búsqueda, en el que el indicador de recálculo se define por $p_i \frac{1 - acc_i}{f_i c_i}$.

6. Un controlador de recálculo (2) para empleo en un entorno de base de datos (1), comprendiendo el entorno de base de datos (1) una plataforma de búsqueda (4) que mantiene resultados de búsqueda precalculados y una plataforma de cálculo (3), el control de recálculo (2) **caracterizado por** estar dispuesto para determinar resultados de búsqueda precalculados para recálculo estando dispuesto para:

- seleccionar de manera iterativa un número de los resultados de búsqueda precalculados para recálculo por la plataforma de cálculo durante un intervalo de tiempo dado que forma un ciclo de recálculo, el número de resultados de búsqueda precalculados estando limitado por los recursos de cálculo disponibles de la plataforma de cálculo durante el intervalo de tiempo dado,

o en el que los recursos de cálculo c_i necesarios para recalcular un resultado de búsqueda precalculado i dependen de si se recalculan o no otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i durante el intervalo de tiempo dado y en el que

o los recursos de cálculo c_i necesarios para recalcular el resultado de búsqueda precalculado i se estiman dinámicamente y de manera iterativa dependiendo de qué otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i se seleccionan para recálculo durante el intervalo de tiempo dado;

en el que seleccionar de manera iterativa un número de los resultados de búsqueda precalculados a recalcularse por la plataforma de cálculo (3) durante un intervalo de tiempo dado y estimar dinámicamente y de manera iterativa los recursos de cálculo c_i necesarios para recalcular el resultado de búsqueda precalculado i durante el intervalo de tiempo dado comprende:

a) inicializar los recursos de cálculo c_i necesarios para recalcular cada resultado de búsqueda precalculado i con un valor que supone el cálculo de i independiente del cálculo de otros resultados de búsqueda precalculados seleccionados para recálculo durante el intervalo de tiempo dado;

b) en una primera iteración de selección, seleccionar una primera porción de los resultados de búsqueda precalculados para recálculo, en el que - basándose en los valores actuales para los recursos de cálculo necesarios para recalcular la porción de resultados de búsqueda precalculados - la primera porción seleccionada requiere un porcentaje por debajo del 100 % de los recursos de cálculo globales de la plataforma de cálculo disponibles para recálculo en el intervalo de tiempo dado;

c) reevaluar los recursos de cálculo c_i necesarios para recalcular cada resultado de búsqueda precalculado i teniendo en cuenta qué resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i se han seleccionado para recálculo;

d) realizar iteraciones adicionales de selección de porciones de resultados de búsqueda precalculados aún no seleccionados adicionales y reevaluar los recursos de cálculo c_i necesarios para recalcular cada resultado de búsqueda precalculado i si se agota menos del 100 % de los recursos de cálculo globales de la plataforma de cálculo (3) disponibles para recálculo en el intervalo de tiempo dado;

- generar y transmitir órdenes de recálculo a la plataforma de cálculo (3) para recalcular el número de resultados de búsqueda precalculados.

7. El controlador de recálculo de la reivindicación 6, que está dispuesto adicionalmente para proporcionar un valor perfeccionado del porcentaje de los recursos de cálculo globales de la plataforma de cálculo disponibles para recálculo en el intervalo de tiempo dado necesarios para recalcular la porción seleccionada de los resultados de búsqueda precalculados en el intervalo de tiempo dado en la función de reevaluar los recursos de cálculo c_i necesarios para recalcular cada resultado de búsqueda precalculado i que incluye.

8. El controlador de recálculo de la reivindicación 6 o la reivindicación 7, el controlador de recálculo (2) que está dispuesto adicionalmente para asignar un indicador de recálculo a cualquiera de los resultados de búsqueda precalculados, en el que el indicador de recálculo para el resultado de búsqueda precalculado i indica una necesidad para recálculo del resultado de búsqueda precalculado i durante el intervalo de tiempo dado y es basándose en al menos los siguientes factores:

o una probabilidad acc_i de que el resultado de búsqueda precalculado i sea aún válido;

o una frecuencia de recálculo f_i del resultado de búsqueda precalculado i aproximada por la recíproca de un

tiempo desde el último recálculo del resultado de búsqueda precalculado i ; y

o una medida para los recursos de cálculo estimados c_i necesarios para recalculer el resultado de búsqueda precalculado i ;

5 - en el que el controlador de recálculo (2) está dispuesto para seleccionar la primera porción y las porciones de los resultados de búsqueda precalculados aún no seleccionados para recálculo de acuerdo con el indicador de recálculo.

9. El controlador de recálculo de la reivindicación 8, en el que el indicador de recálculo se define por $\frac{1 - acc_i}{f_i c_i}$.

10. El controlador de recálculo de la reivindicación 8 o la reivindicación 9, en el que el indicador de recálculo está basado adicionalmente en una frecuencia de acceso p_i del resultado de búsqueda precalculado i desde la plataforma

de búsqueda, en el que el indicador de recálculo se define por $p_i \frac{1 - acc_i}{f_i c_i}$.

11. Un programa informático que cuando se ejecuta en un sistema informático provoca que el sistema informático determine resultados de búsqueda precalculados para recálculo:

seleccionar de manera iterativa un número de los resultados de búsqueda precalculados para recálculo por una plataforma de cálculo durante un intervalo de tiempo dado que forma un ciclo de recálculo, el número de resultados de búsqueda precalculados estando limitado por los recursos de cálculo de la plataforma de cálculo disponibles durante el intervalo de tiempo dado,

- en el que los recursos de cálculo c_i necesarios para recalculer un resultado de búsqueda precalculado i dependen de si se recalculan o no otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i durante el intervalo de tiempo dado y en el que

- los recursos de cálculo c_i necesarios para recalculer el resultado de búsqueda precalculado i se estiman dinámicamente y de manera iterativa dependiendo de qué otros resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i se seleccionan para recálculo durante el intervalo de tiempo dado;

en el que seleccionar de manera iterativa un número de los resultados de búsqueda precalculados a recalcularse por la plataforma de cálculo (3) durante un intervalo de tiempo dado y estimar dinámicamente y de manera iterativa los recursos de cálculo c_i necesarios para recalculer resultado de búsqueda precalculado i durante el intervalo de tiempo dado comprende:

a) inicializar los recursos de cálculo c_i necesarios para recalculer cada resultado de búsqueda precalculado i con un valor que supone el cálculo de i independiente del cálculo de otros resultados de búsqueda precalculados seleccionados para recálculo durante el intervalo de tiempo dado;

b) en una primera iteración, seleccionar una primera porción de los resultados de búsqueda precalculados para recálculo, en el que - basándose en los valores actuales para los recursos de cálculo necesarios para recalculer la porción de resultados de búsqueda precalculados - la porción seleccionada requiere un porcentaje por debajo del 100 % de los recursos de cálculo globales de la plataforma de cálculo disponibles para recálculo en el intervalo de tiempo dado;

c) reevaluar los recursos de cálculo c_i necesarios para recalculer cada resultado de búsqueda precalculado i teniendo en cuenta qué resultados de búsqueda precalculados relacionados con el resultado de búsqueda precalculado i se han seleccionado para recálculo;

d) realizar iteraciones adicionales de selección de porciones de resultados de búsqueda precalculados aún no seleccionados adicionales y reevaluar los recursos de cálculo c_i necesarios para recalculer cada resultado de búsqueda precalculado i si se agota menos del 100 % de los recursos de cálculo globales de la plataforma de cálculo (3) disponibles para recálculo en el intervalo de tiempo dado; generar y transmitir órdenes de recálculo a la plataforma de cálculo (3) para recalculer el número de resultados de búsqueda precalculados.

12. El programa informático de la reivindicación 11, que está dispuesto adicionalmente para provocar que el sistema informático proporcione un valor perfeccionado del porcentaje de los recursos de cálculo globales de la plataforma de cálculo disponibles para recálculo en el intervalo de tiempo dado necesarios para recalculer la porción seleccionada de los resultados de búsqueda precalculados en el intervalo de tiempo dado en la función de reevaluar los recursos de cálculo c_i necesarios para recalculer cada resultado de búsqueda precalculado i que incluye.

13. El programa informático de la reivindicación 11 o 12, que provoca que el sistema informático asigne un indicador de recálculo a cualquiera de los resultados de búsqueda precalculados, en el que el indicador de recálculo para el resultado de búsqueda precalculado i indica una necesidad para recálculo del resultado de búsqueda precalculado i durante el intervalo de tiempo dado y está basado en al menos los siguientes factores:

- 5
- una probabilidad acc_i de que el resultado de búsqueda precalculado i sea aún válido;
 - una frecuencia de recálculo f_i del resultado de búsqueda precalculado i aproximada por la recíproca de un tiempo desde el último recálculo del resultado de búsqueda precalculado i ; y
 - una medida para los recursos de cálculo estimados c_i necesarios para recalcular el resultado de búsqueda precalculado i ;
- 10

- en el que el programa informático provoca que el sistema informático seleccione la primera porción y las porciones de los resultados de búsqueda precalculados aún no seleccionados para recálculo de acuerdo con el indicador de recálculo.

15

14. El programa informático de la reivindicación 13, en el que el indicador de recálculo se define por $\frac{1 - acc_i}{f_i c_i}$.

15. El programa informático de la reivindicación 13 o la reivindicación 14, en el que el indicador de recálculo está basado adicionalmente en una frecuencia de acceso p_i del resultado de búsqueda precalculado i desde la plataforma

de búsqueda, en el que el indicador de recálculo se define por $p_i \frac{1 - acc_i}{f_i c_i}$.

20

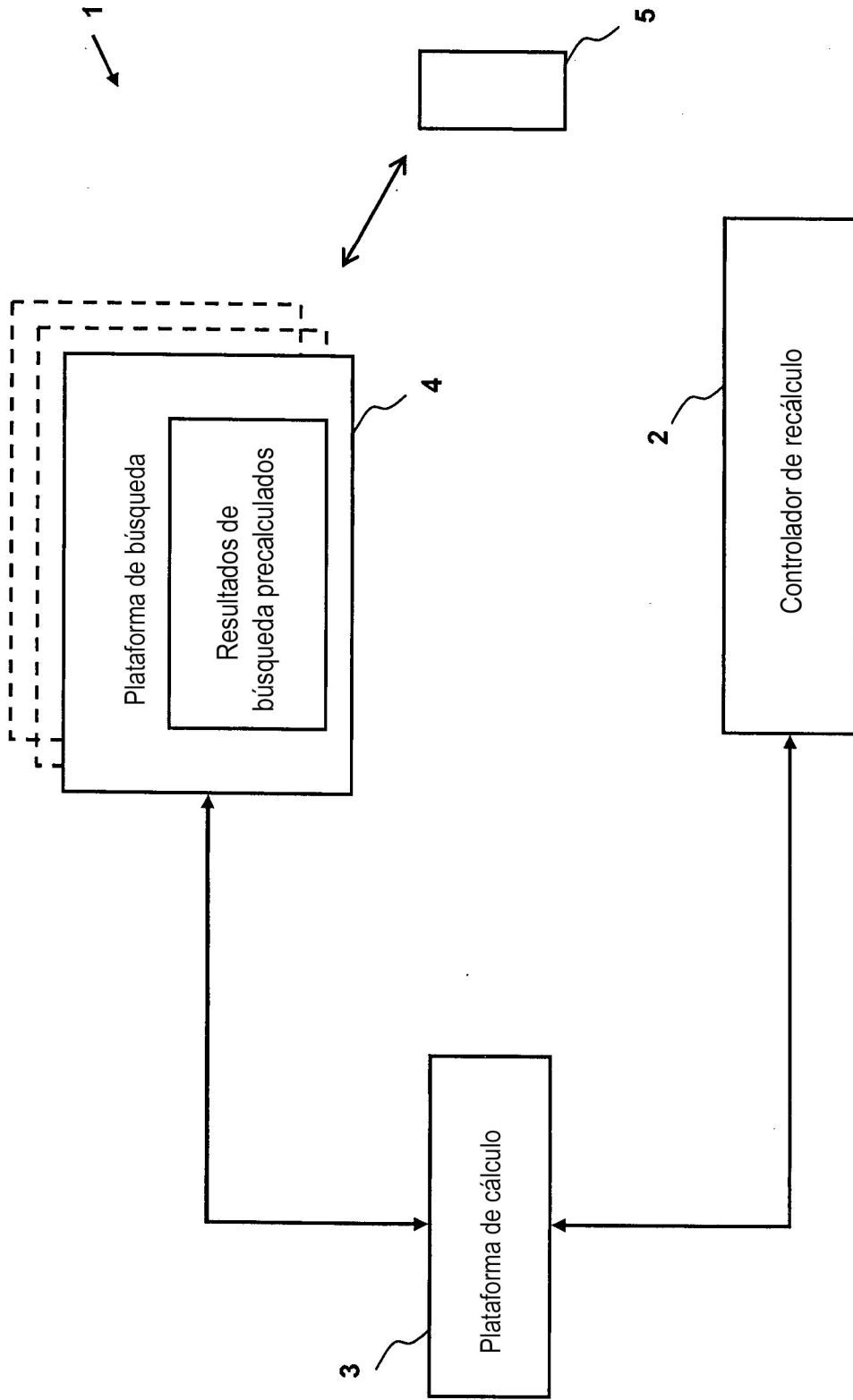


Fig. 1

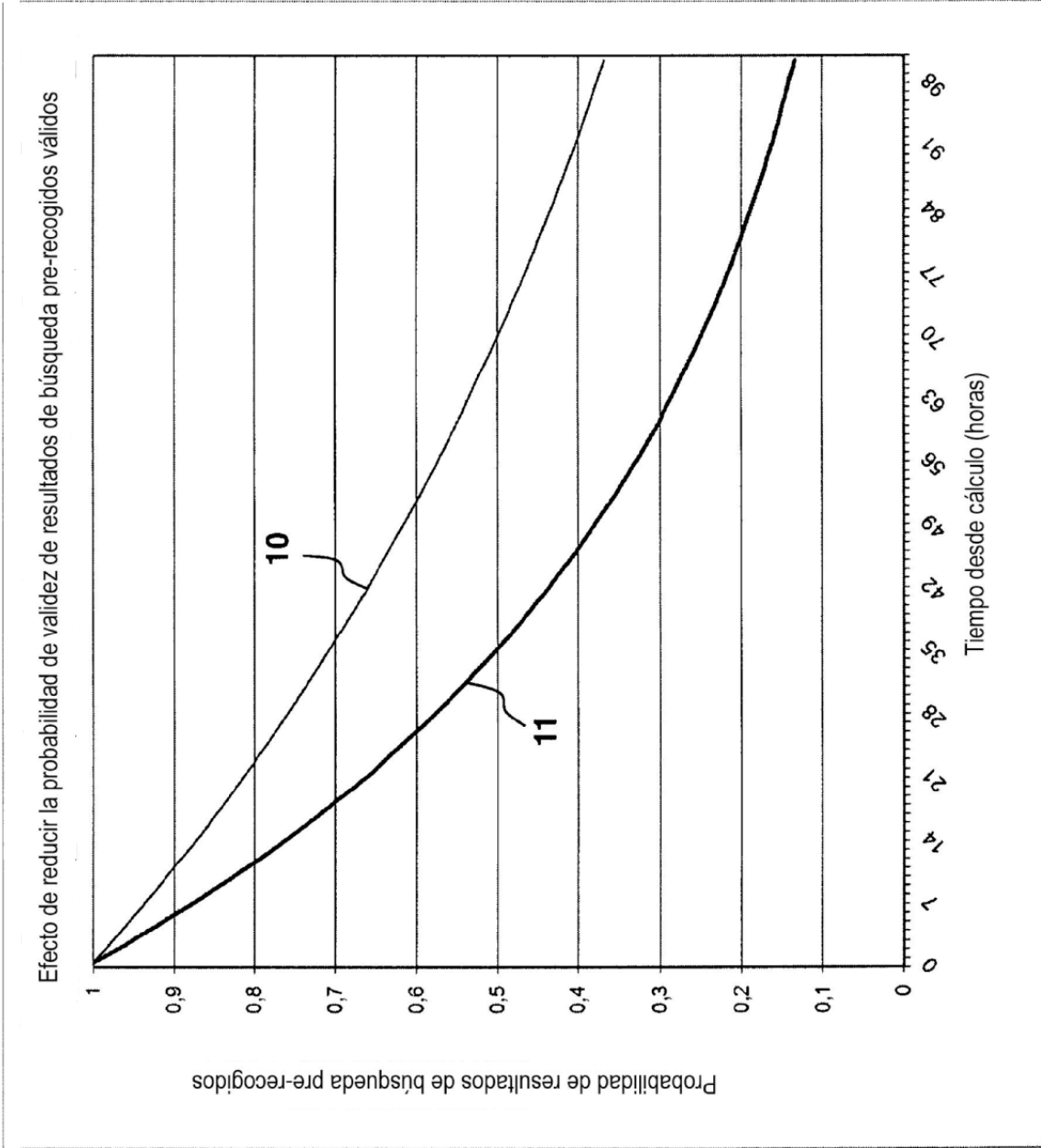


Fig. 2

Tendencia de precisión global

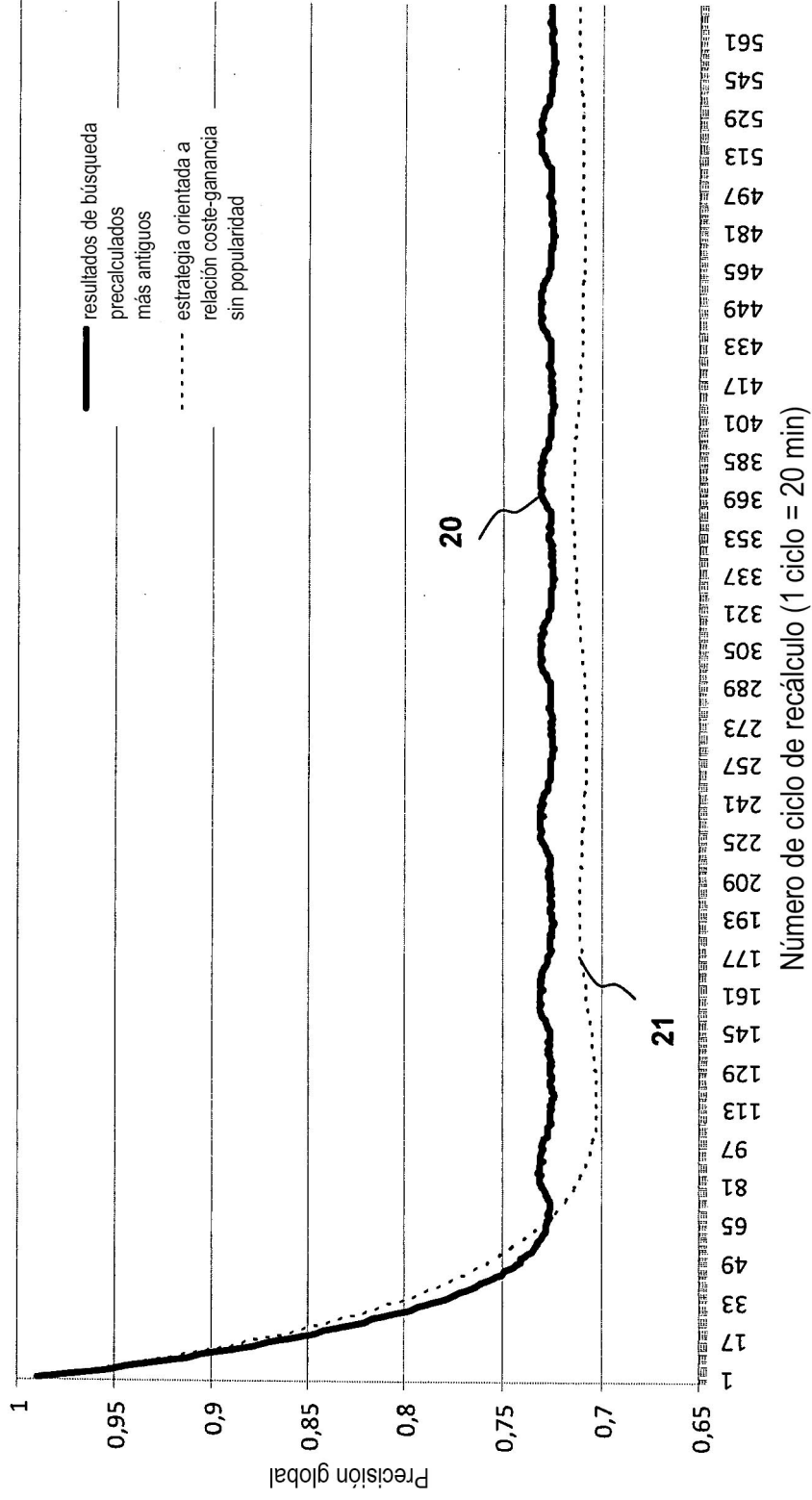


Fig. 3a

Tendencia de precisión de usuario

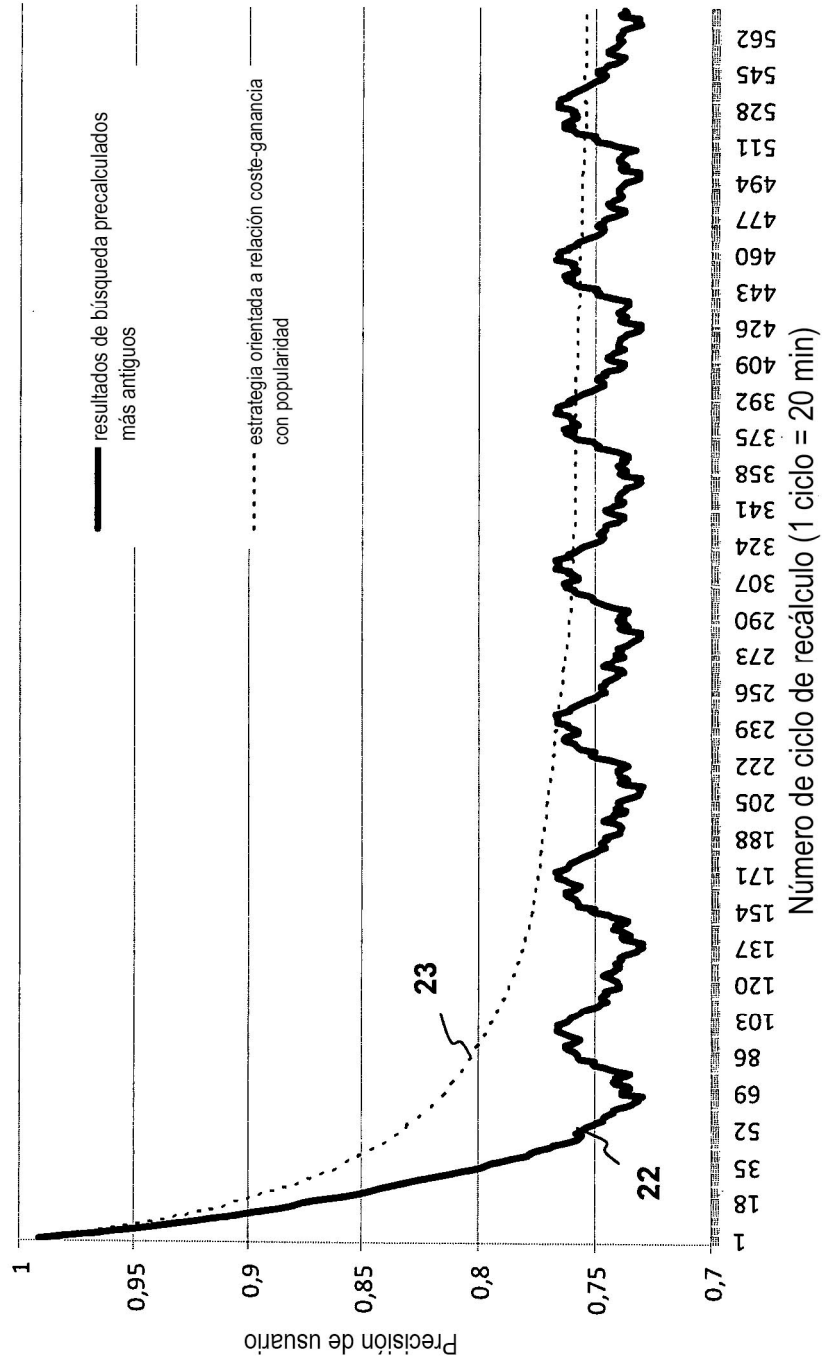


Fig. 3b

Distribución de precisión (acumulativa)

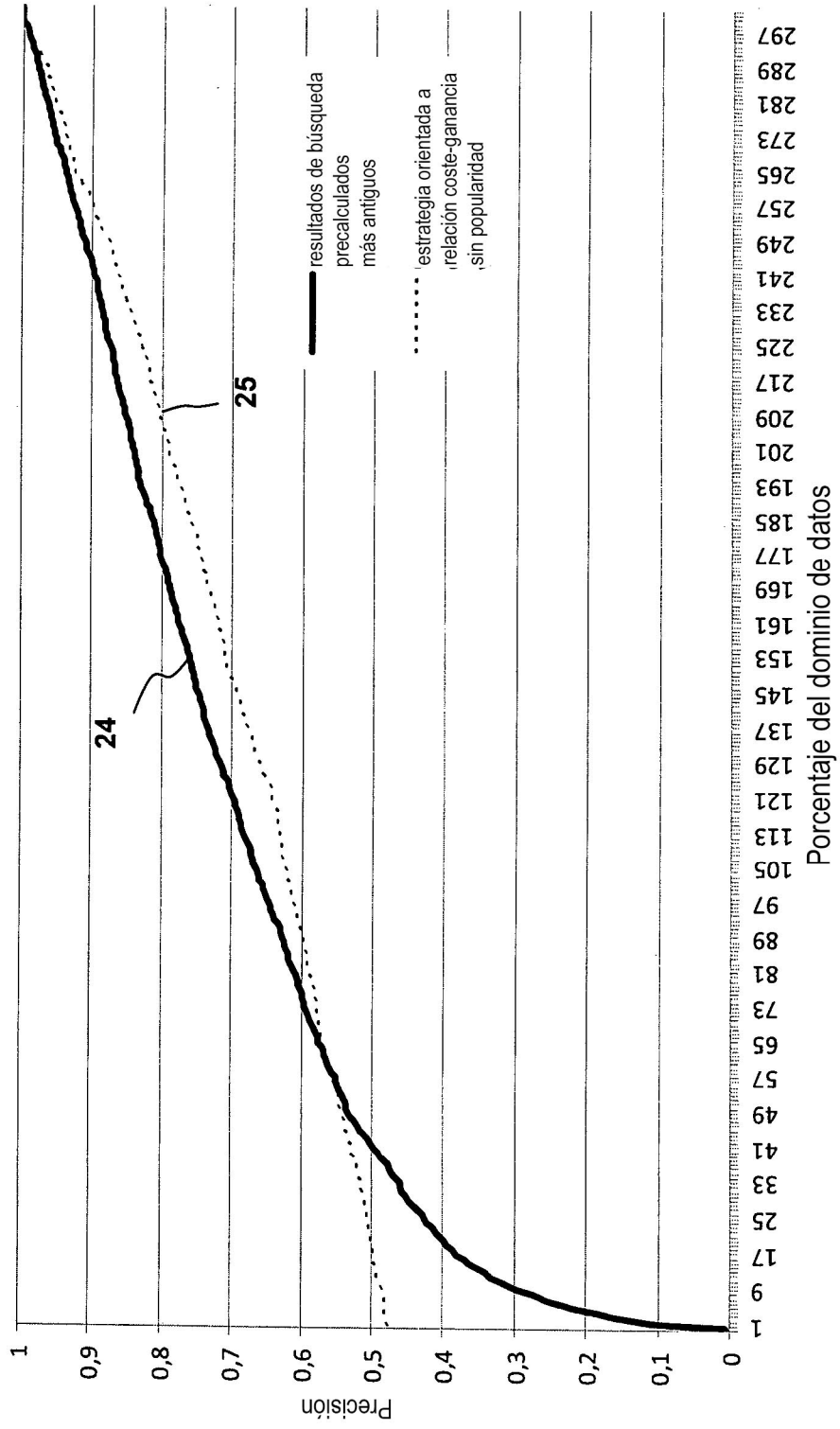


Fig. 3c

Tendencia de precisión de usuario

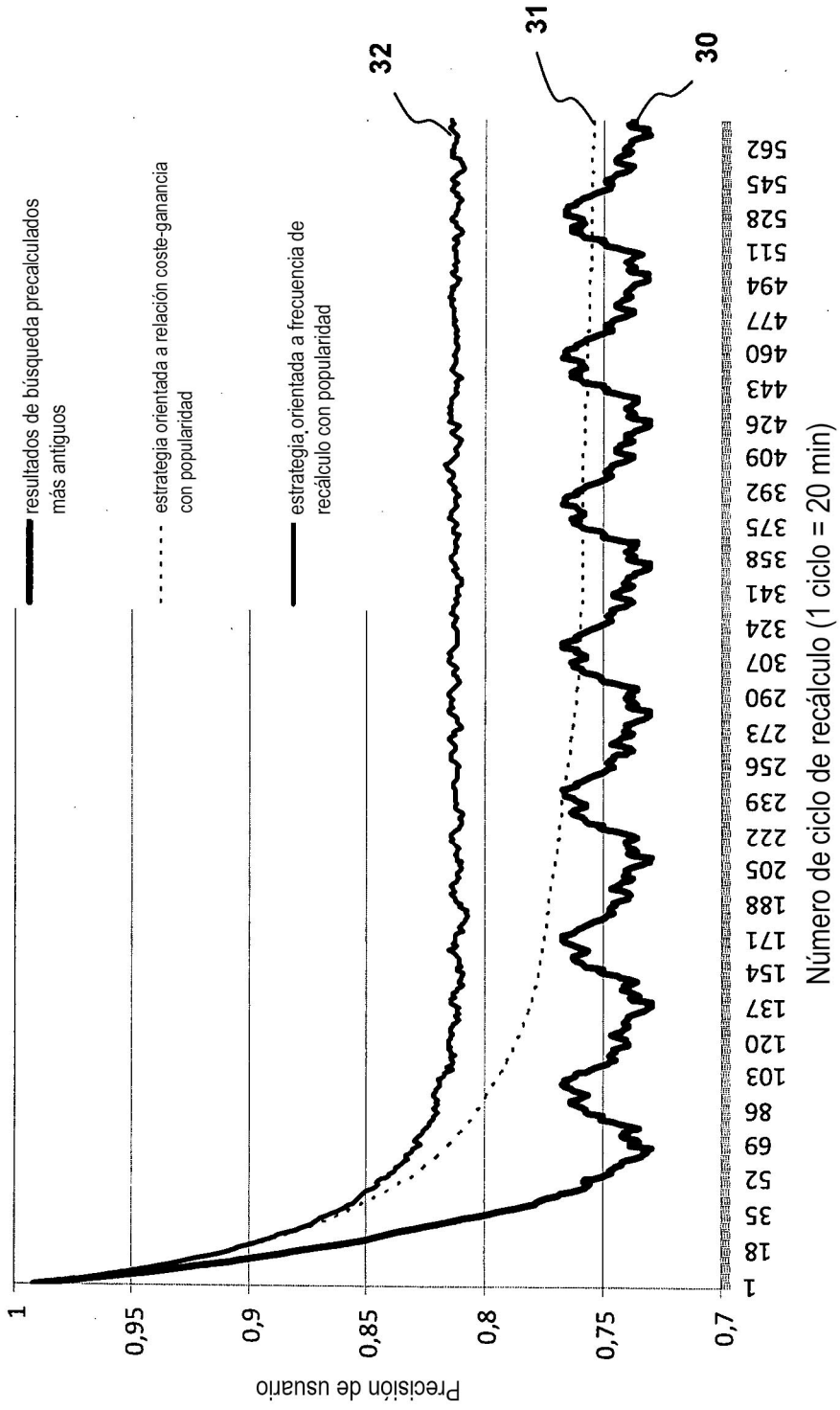


Fig. 4a

Tendencia de precisión global

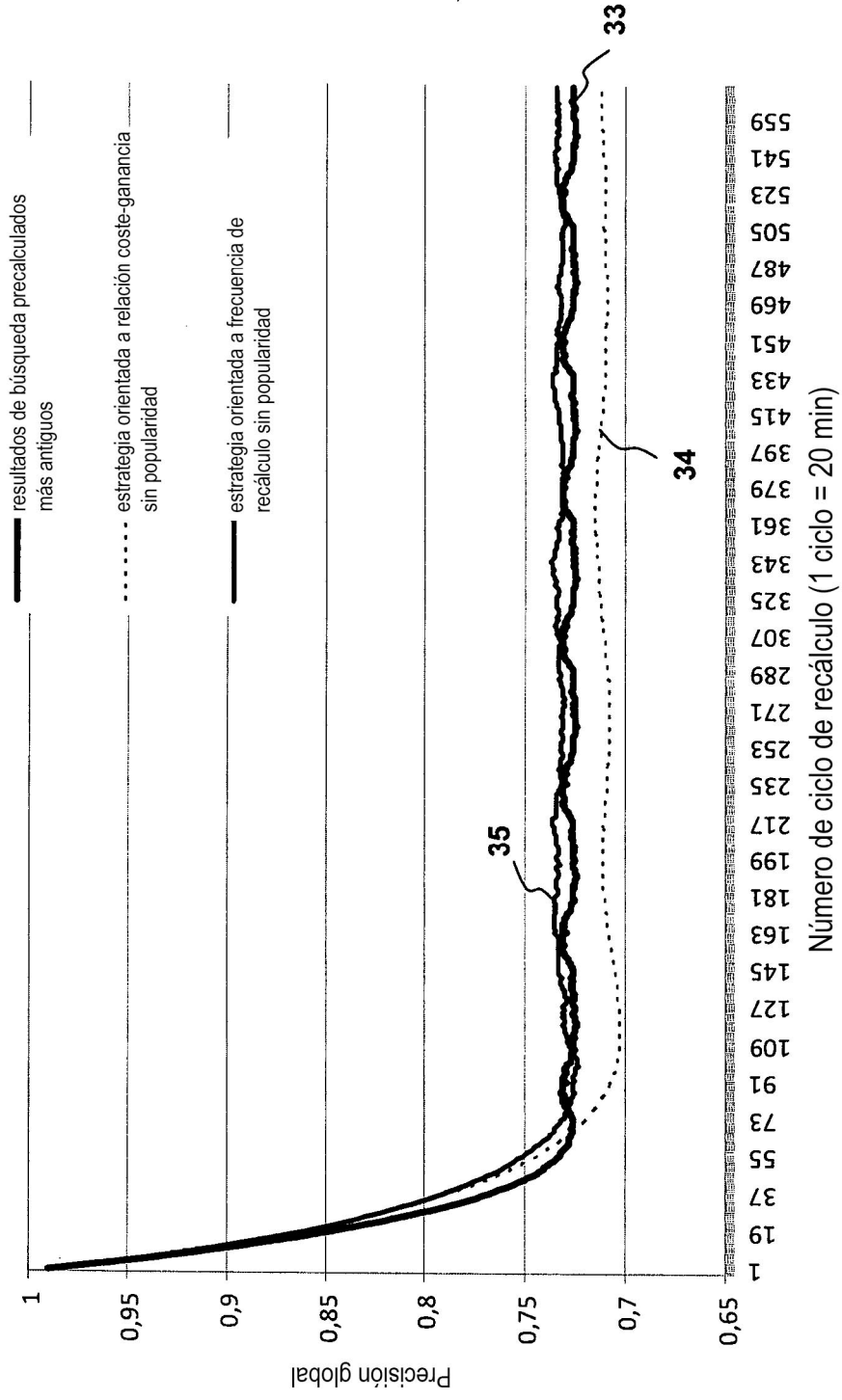


Fig. 4b

Recursos de cálculo en dependencia de número de registros de vuelo precalculados

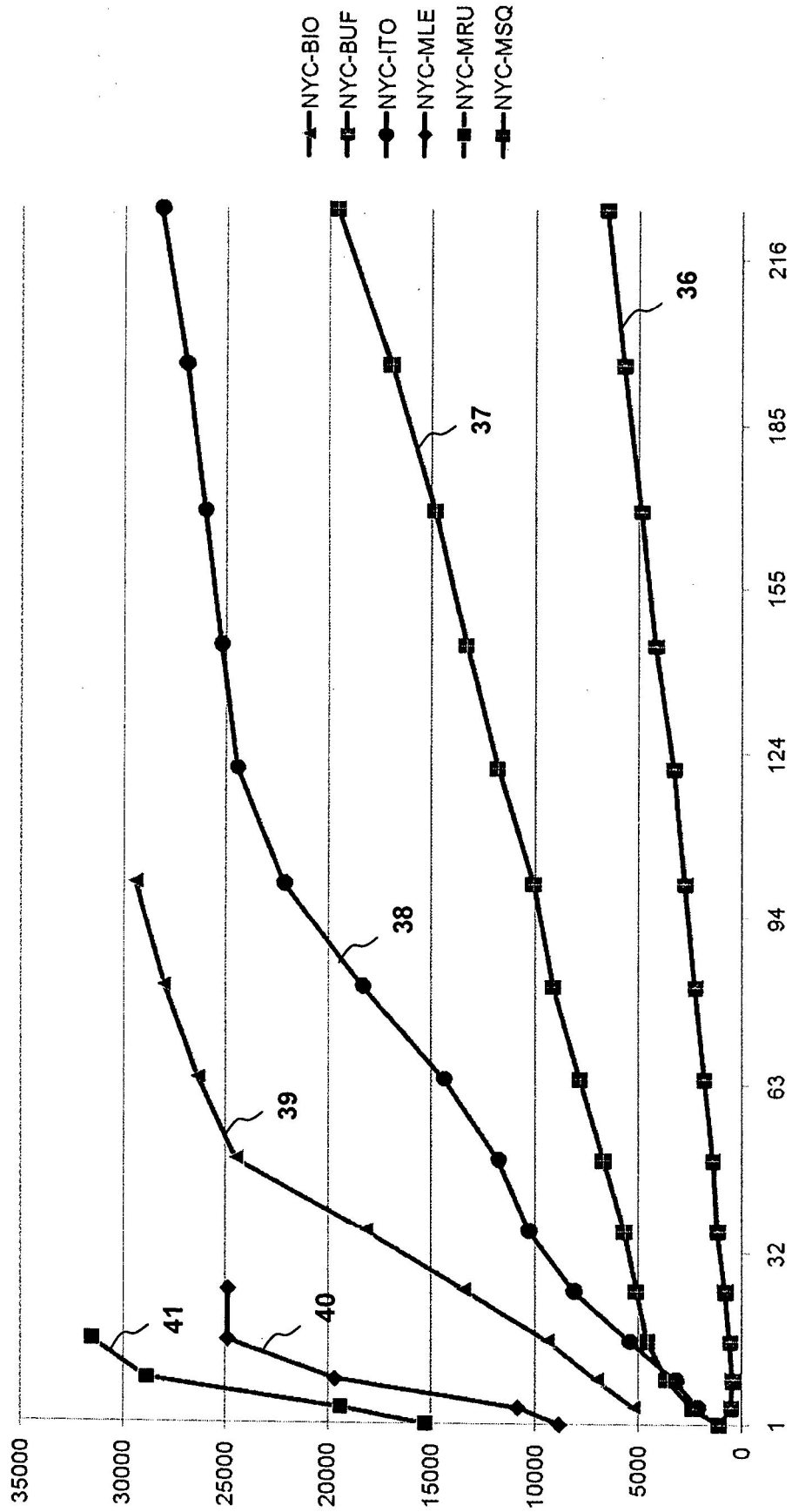


Fig. 5

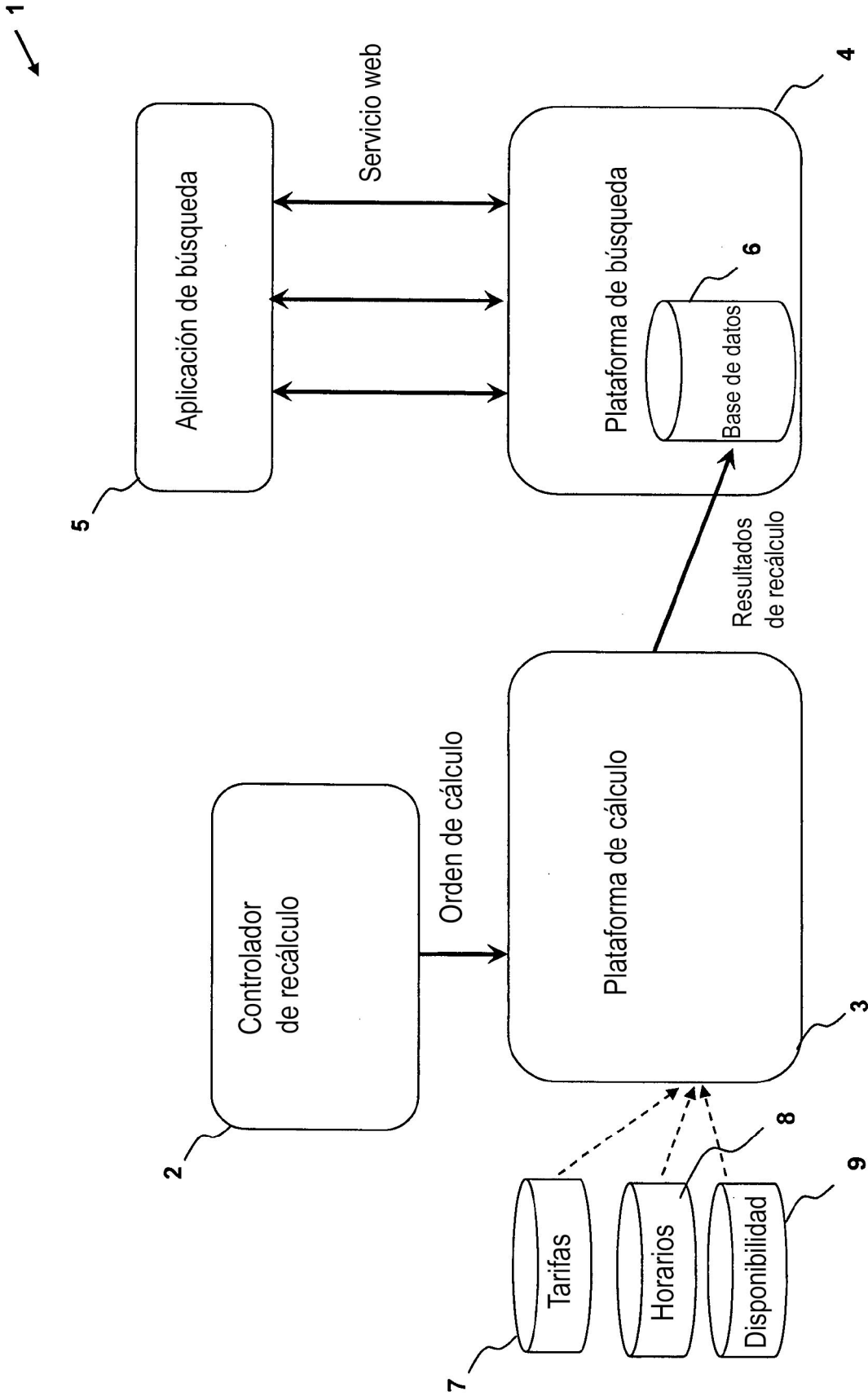


Fig. 6

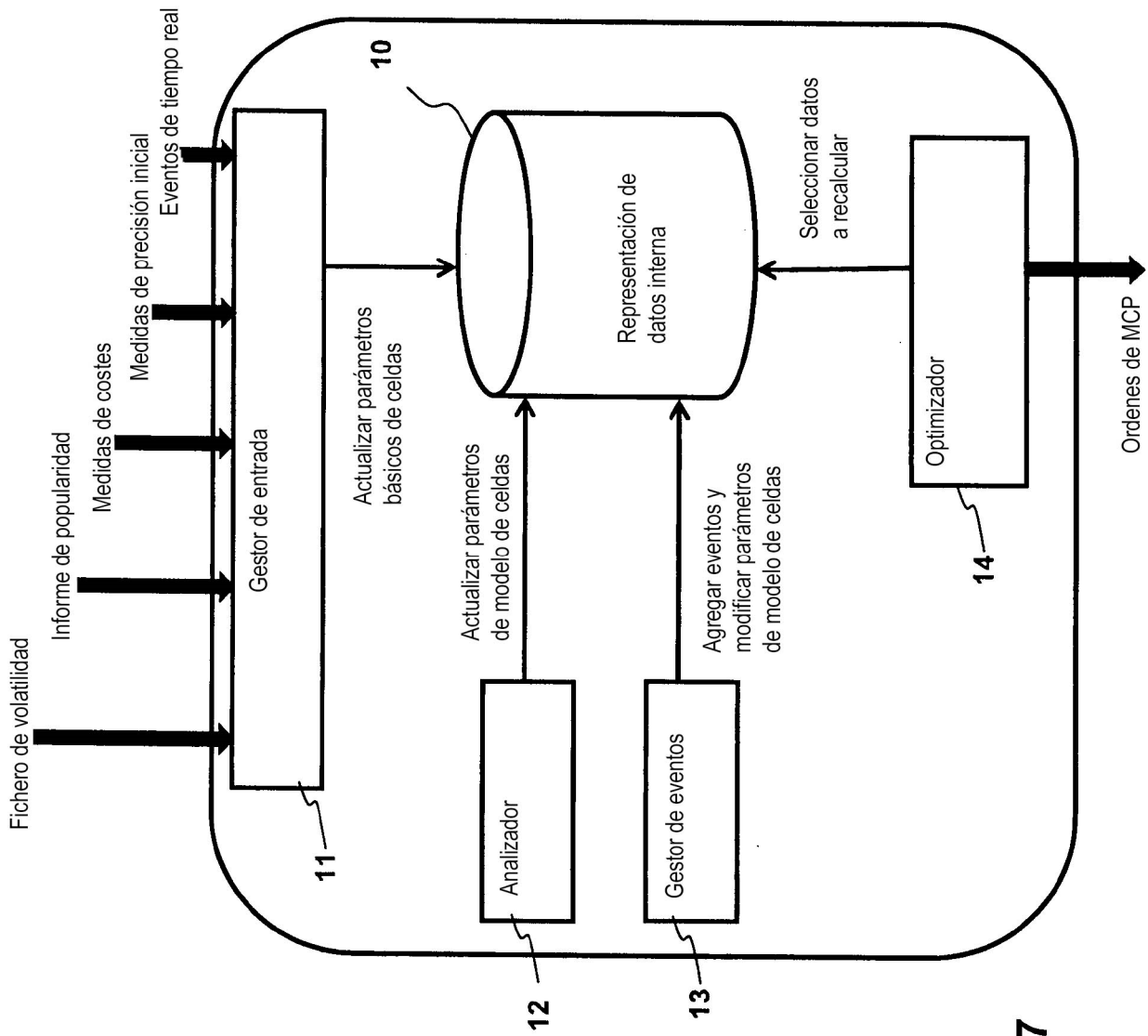


Fig. 7

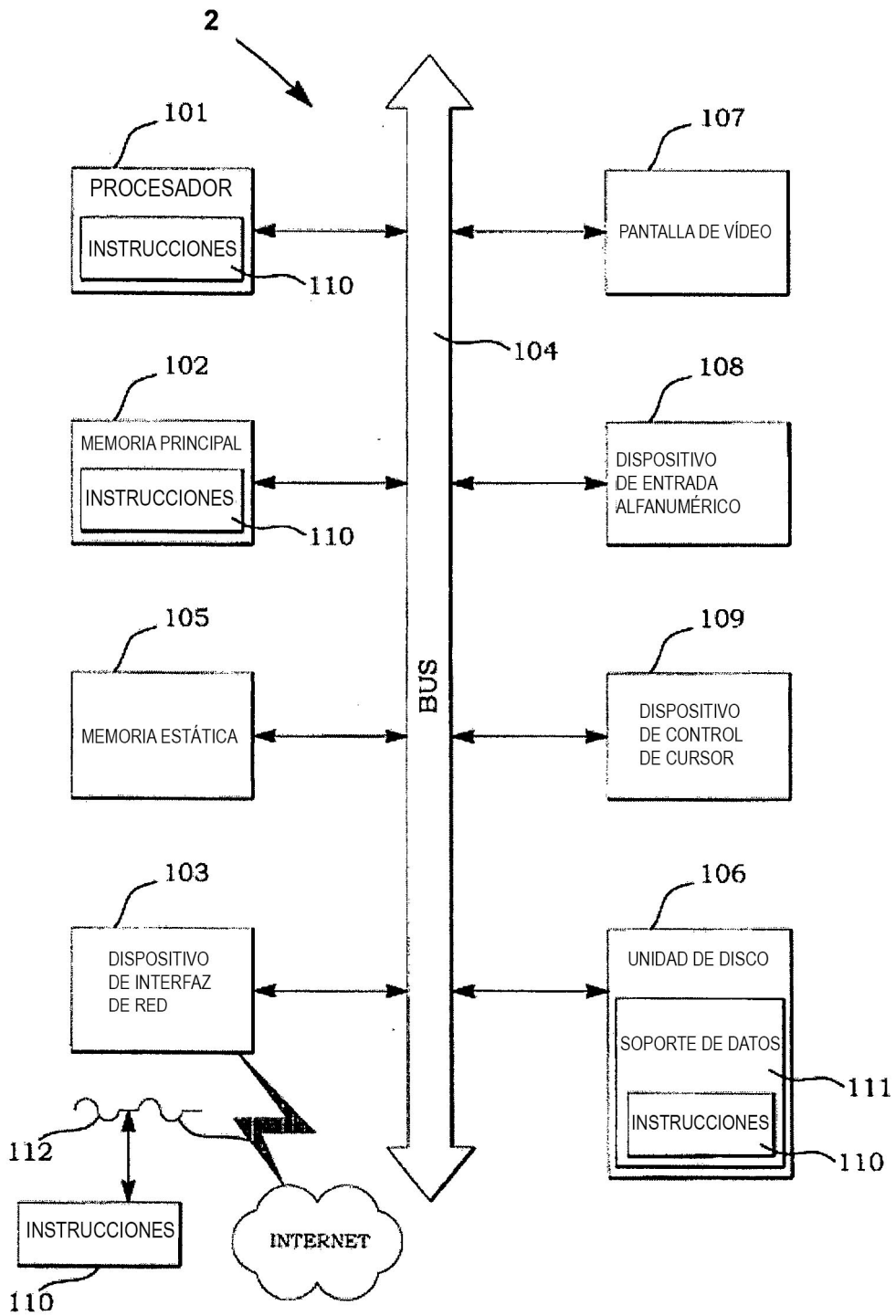


Fig. 8