

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 990 104**

51 Int. Cl.:

G06F 16/2455 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.11.2014** **E 14003698 (9)**

97 Fecha y número de publicación de la concesión europea: **31.07.2024** **EP 3016000**

54 Título: **Gestión de resultados de búsqueda precalculados**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
28.11.2024

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:

ELZABURU, S.L.P

ES 2 990 104 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

Gestión de resultados de búsqueda precalculados

Campo de la invención

5 La presente invención se refiere, en general, a la tecnología de la información. Más específicamente, la presente invención se refiere a la tecnología de base de datos y a mecanismos para recalcular resultados de búsqueda precalculados.

Antecedentes

10 Se conocen el precálculo de resultados de búsqueda y la devolución de resultados de búsqueda precalculados en respuesta a consultas de búsqueda en lugar de solo calcular los resultados de búsqueda en el momento de la consulta. De esta manera, se pueden acortar los tiempos de respuesta. Los resultados de búsqueda precalculados deben mantenerse actualizados para garantizar que se devuelvan respuestas válidas. Si los datos subyacentes (que son la base para el precálculo de los resultados de búsqueda) cambian, los resultados de búsqueda precalculados pueden desactualizarse y se devolverán resultados incorrectos. Por lo tanto, se emplean estrategias de actualización para mantener los resultados de búsqueda precalculados actualizados.

15 En la técnica anterior se conocen diversas estrategias de actualización relativamente simples. Por ejemplo, el documento WO 01/33472 se refiere a un sistema de disponibilidad usado en un sistema de planificación de viajes. El sistema incluye una memoria caché que tiene entradas de información de disponibilidad con respecto a asientos de aerolínea. Un gestor de memoria caché gestiona la información de entrada en la memoria caché para mantener la información en la memoria caché correcta, actual, completa o de otro modo lo más útil posible. En respuesta a una consulta dirigida a la memoria caché, el gestor de memoria 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 la memoria caché a modificar se obtienen mediante notificaciones asíncronas de sistemas externos y se determinan mediante un modelo determinista, predictivo o estadístico.

20 De manera similar, el documento WO 02/25557 pertenece a un sistema de recuperación de información en donde la información recibida de fuentes de información se almacena en caché para uso futuro como, por ejemplo, para futuras solicitudes de clientes. Pueden generarse consultas proactivas para rellenar una memoria caché y/o para actualizar información actualmente almacenada en memoria caché. En un sistema de información de aerolíneas, se ordenan consultas proactivas basándose en estadísticas o indicaciones predictivas como, por ejemplo, cercanía de la hora de partida, antigüedad de datos almacenados en caché, asientos restantes en una aeronave, vacaciones o eventos especiales o tipo de equipo. Además, las actualizaciones son recibidas por notificaciones externas de aerolíneas como, por ejemplo, mensajes AVS.

25 Además, el documento WO 99/22315 describe un mecanismo para actualizar automáticamente documentos en una memoria caché usando un modelo probabilístico basado en estadísticas. Para cada documento, la memoria caché determina una probabilidad $\Psi_i(t)$ de que un objeto almacenado en caché i sea obsoleto en un tiempo particular t (es decir, el servidor ha cambiado ese objeto) y una probabilidad $P_i(h)$ de que un usuario solicite el objeto i en el tiempo de solicitud h . La memoria caché actualiza esos objetos con el producto más alto $P_i = \Psi_i(t) \times P_i(h)$, es decir, la probabilidad de que un objeto desactualizado se devuelva al usuario con la siguiente solicitud. Para mantener estos valores de probabilidad, la memoria caché mantiene y sigue estadísticas históricas para los objetos almacenados en memoria caché como, por ejemplo, un intervalo medio estimado entre actualizaciones de servidor EUI. La EUI de un objeto se actualiza, p. ej., cuando el propio objeto es actualizado por el servidor o el objeto no se actualiza después de que haya transcurrido su tiempo medio estimado de actualización.

30 Además, los documentos WO 2014/026753 A1 y WO 2013/160721 A1 del solicitante describen posibles implementaciones de un modelo probabilístico para aproximar probabilidades de validez de resultados de búsqueda precalculados. Además, se proponen mecanismos para reconocer eventos asíncronos en tiempo real que no están representados en el modelo probabilístico y mecanismos para considerar el impacto potencial de dichos eventos en tiempo real en la validez de los resultados de búsqueda calculados previamente.

Compendio de la invención

La presente invención se define por las reivindicaciones independientes.

35 Según un primer aspecto, se provee un método de gestión de resultados de búsqueda precalculados. El método se lleva a cabo en un entorno de base de datos. El entorno de base de datos incluye al menos una plataforma de búsqueda, un controlador de recálculo y una plataforma de cálculo. La plataforma de búsqueda mantiene resultados de búsqueda calculados previamente. Los resultados de búsqueda precalculados se subdividen en múltiples partes de resultados de búsqueda precalculados relacionados que incluyen una primera parte D. El controlador de recálculo controla el recálculo de los resultados de búsqueda precalculados basándose en probabilidades de validez que se asocian a los resultados de búsqueda precalculados. La plataforma de cálculo recalcula los resultados de búsqueda precalculados. El controlador de recálculo determina una tendencia de validez λ_i que indica una tasa de cambio de la

probabilidad de validez de un resultado de búsqueda i precalculado en el tiempo. El resultado de búsqueda i precalculado es un miembro de la primera parte D. La tendencia de validez λ_i se deriva de al menos tres recálculos anteriores de i . El controlador de recálculo determina además una tasa de validez instantánea $\hat{\lambda}_i$ para el resultado i de búsqueda precalculado. La tasa de validez instantánea $\hat{\lambda}_i$ se deriva de los dos últimos recálculos de los resultados de búsqueda precalculados relacionados incluidos en la primera parte D. En respuesta a la determinación de que una diferencia entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i supera un grado dado, el controlador de recálculo disminuye las probabilidades de validez que se asocian a otros resultados de búsqueda precalculados incluidos en una segunda parte D' que se correlaciona con la primera parte D. La cantidad de disminución depende de la cantidad de la diferencia entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i . Finalmente, el controlador de recálculo emite órdenes de recálculo a la plataforma de cálculo para recalcular una parte de los resultados de búsqueda precalculados, seleccionándose la parte por el controlador de recálculo basándose en las probabilidades de validez asociadas a los resultados de búsqueda precalculados, a saber, resultados de búsqueda precalculados con probabilidades de validez más bajas que otros resultados de búsqueda precalculados con probabilidades de validez más altas.

Según otro aspecto, se provee un controlador de recálculo equipado con una funcionalidad respectiva.

Según incluso otro aspecto, se provee un programa informático que, cuando se ejecuta en un sistema informático, lleva a cabo el método como se ha descrito más arriba. El programa informático puede almacenarse como instrucciones ejecutables en un medio de almacenamiento legible por ordenador no transitorio.

Se exponen aspectos adicionales en las reivindicaciones dependientes.

Breve descripción de las figuras

La presente invención se describirá con referencia a las figuras anexas. Los números de referencia similares indican generalmente elementos idénticos o funcionalmente similares.

La Fig. 1 muestra esquemáticamente la arquitectura básica de un entorno de base de datos.

La Fig. 2 ilustra el efecto de la probabilidad de validez decreciente de resultados de búsqueda precalculados a lo largo del tiempo.

La Fig. 3 presenta una vista más detallada de la estructura de los resultados de búsqueda precalculados mantenidos por la plataforma de búsqueda.

Las Fig. 4a y 4b visualizan intervalos de tiempo inequidistantes entre varios recálculos sucesivos de un resultado de búsqueda precalculado.

Las Fig. 5a y 5b visualizan intervalos de tiempo inequidistantes entre dos recálculos de múltiples resultados de búsqueda precalculados relacionados.

La Fig. 6 es un diagrama de flujo que representa el proceso de detección de un evento perjudicial para la validez de los resultados de búsqueda calculados previamente a un nivel alto.

La Fig. 7 es un diagrama de flujo que representa un enfoque para determinar una tendencia de validez y/o una tasa de validez instantánea estableciendo una función de distribución empírica.

Las Fig. 8a y 8b visualizan el establecimiento de una función de distribución empírica utilizando períodos de estabilidad que abarcan un recálculo de un resultado de búsqueda precalculado y una ventana que se desliza durante un período de estabilidad.

Las Fig. 9a y 9b muestran los valores de probabilidad de validez discretos resultantes y la obtención de la tendencia de validez/tasa de validez instantánea por regresión.

La Fig. 10 muestra un ejemplo de resultados de búsqueda precalculados con probabilidades de validez asociadas.

La Fig. 11 visualiza la adaptación de la probabilidad de validez modelada de un resultado de búsqueda precalculado después de haber detectado una diferencia sustancial entre la tendencia de validez y la tasa de validez instantánea de otro resultado de búsqueda precalculado correlacionado.

La Fig. 12 representa una estructura interna a modo de ejemplo de un controlador de recálculo.

La Fig. 13 es una vista esquemática a modo de ejemplo de la arquitectura interna de un ordenador/servidor que implementa una configuración como se describe en la presente memoria.

Descripción detallada

Antes de pasar a la descripción detallada con referencia a las Fig. 7 a 13, se expondrán en primer lugar algunos aspectos más generales con referencia a las Fig. 1 a 6.

- 5 Como ya se ha indicado al principio, las metodologías descritas en la presente memoria se refieren a sistemas de base de datos que ofrecen resultados de búsqueda calculados previamente a clientes. El objetivo técnico del pre cálculo es generalmente disminuir los tiempos de respuesta para responder a consultas de búsqueda. En lo sucesivo, los términos "pre cálculo" y "precalculado" se usan para cubrir cualquier tipo de pre cálculo y prerrecogida como, por ejemplo, rastreadores de Internet que recopilan o copian el contenido de servidores web de Internet, pero también cálculos complejos e intensivos en tiempo de resultados de búsqueda sobre la base de datos subyacentes como se describe, p. ej., para recomendaciones de viaje con precios por los documentos PCT/EP2013/002390 y EP 10 2541473 A1. El término "base de datos" pretende abarcar cualquier tipo de sistema de almacenamiento de información estructurada como, por ejemplo, bases de datos autónomas estándar como el servidor SQL o bases de datos Oracle, así como sistemas de almacenamiento complejos, distribuidos y/o patentados, bases de datos relacionales que incluyen sistemas de gestión de bases de datos o sistemas de bases de datos orientados a objetos y similares.
- 15 La arquitectura de un sistema 1 de base de datos a modo de ejemplo se muestra en la Fig. 1. El sistema 1 de base de datos incluye al menos uno, pero generalmente múltiples clientes 5 y al menos una plataforma 4 de búsqueda. Para aumentar la seguridad y/o el rendimiento de las fallas, puede haber múltiples plataformas 4 de búsqueda. La al menos una plataforma 4 de búsqueda mantiene resultados de búsqueda calculados previamente con el fin de disminuir los tiempos de respuesta para responder consultas de búsqueda recibidas de los clientes 5.
- 20 El cliente 5 dirige consultas de búsqueda a la plataforma 4 de búsqueda, incluyendo cada consulta de búsqueda uno o más criterios de búsqueda que restringen la búsqueda. Por ejemplo, si una consulta de búsqueda es una búsqueda de Internet, la consulta de búsqueda puede llevar una cadena de búsqueda, texto de búsqueda o frase de búsqueda como criterios de búsqueda. Otro criterio de búsqueda puede ser el lenguaje de sitios web que se van a buscar o una indicación de un momento de la primera disponibilidad de la cadena de búsqueda solicitada, texto de búsqueda o frase de búsqueda. Según otro ejemplo, la consulta de búsqueda es una solicitud de base de datos para un producto o 25 servicio ofrecido por una plataforma de proveedores de servicios como, por ejemplo, un almacén de libros de Internet o un proveedor de viajes. En ese caso, la consulta de búsqueda puede incluir, p. ej., un límite de precio superior o un intervalo de precios para el servicio o producto y características deseadas del producto/servicio como, por ejemplo, título de libro, origen y destino de viaje, etc.
- 30 La plataforma 4 de búsqueda procesa una consulta de búsqueda recibida del cliente 5 y lleva a cabo una búsqueda en la base de datos dentro de los resultados de búsqueda calculados previamente. A su vez, la plataforma 4 de búsqueda responde con uno o más resultados de búsqueda precalculados que cumplen los criterios de búsqueda incluidos en la consulta de búsqueda. La manera del procesamiento llevado a cabo por la plataforma 4 de búsqueda no es relevante para las metodologías en la presente memoria, p. ej., si la plataforma 4 de búsqueda recupera solo 35 resultados de búsqueda precalculados que cumplen estrictamente con los criterios de búsqueda incluidos en la consulta de búsqueda o, p. ej., si la plataforma 4 de búsqueda implementa una búsqueda difusa y, por tanto, también devuelve resultados difusos más allá de las restricciones de los criterios de búsqueda en un grado dado. El cliente 5 recibe esta respuesta y procesa la respuesta, p. ej., presenta los resultados de búsqueda al usuario.
- 40 Los resultados de búsqueda precalculados pueden desactualizarse (en lo sucesivo también denominados invalidados) después de un cierto tiempo posterior a su pre cálculo. Generalmente, la causa de invalidación es un cambio en los datos subyacentes u originales. Por ejemplo, el contenido de un servidor web puede cambiar en un cierto punto de tiempo o tarifas subyacentes a las recomendaciones de viaje con precios precalculadas (p. ej., conjuntos de vuelos con precios que forman un viaje) pueden actualizarse de vez en cuando. A partir de estos puntos de tiempo en adelante, los resultados de búsqueda calculados previamente correspondientes almacenados en la plataforma 4 de 45 búsqueda que se ven afectados por el cambio de los datos subyacentes no son válidos. Por lo tanto, los clientes 5 provistos de estos resultados de búsqueda precalculados inválidos recibirán respuestas incorrectas a sus consultas de búsqueda. La solución general de este problema de invalidación es volver a calcular los resultados de búsqueda precalculados de manera regular, irregular o continua.
- 50 El pre cálculo de los resultados de búsqueda se gestiona mediante el controlador 2 de recálculo. El controlador 2 de recálculo puede proveerse como una entidad individual (como se muestra en la Fig. 1) o puede, alternativamente, integrarse en una o (si están presentes) múltiples plataformas 4 de búsqueda. Como los recursos de cálculo para el recálculo son generalmente limitados, los resultados de búsqueda precalculados necesitan priorizarse, es decir, solo una porción de todos los resultados de búsqueda precalculados mantenidos por la plataforma 4 de búsqueda puede recalcularse dentro de un cierto período. Por lo tanto, el controlador 2 de recálculo lleva a cabo una selección de 55 ciertas porciones de los resultados de búsqueda precalculados para el recálculo. Para este fin, el controlador 2 de recálculo gestiona el recálculo de los resultados de búsqueda precalculados según una estrategia de recálculo como, por ejemplo, se describe en el documento WO 2014/026753. Según la estrategia de recálculo empleada, el controlador 2 de recálculo genera y transmite órdenes de recálculo a la plataforma 3 de cálculo, indicando las órdenes de recálculo a la plataforma 3 de cálculo qué resultados de búsqueda precalculados deben recalcularse. En respuesta a la 60 recepción de una orden de recálculo del controlador 2 de recálculo, la plataforma 3 de cálculo ejecuta el recálculo, p.

ej., solicitando datos originales correspondientes a los resultados de búsqueda precalculados de fuentes de datos primarias y/o llevando a cabo un recálculo de los resultados de búsqueda respectivos basándose en datos subyacentes.

5 Ciertos documentos de la técnica anterior como, por ejemplo, los documentos WO 99/22315 y WO 2014/026753 proponen una estrategia de recálculo basada en un modelo probabilístico. En general, tal modelo probabilístico puede incluir, por ejemplo, parámetros como, por ejemplo, una edad, una popularidad (una tasa de acceso al resultado de búsqueda precalculado al consultar a los clientes 5), una tasa de disminución de una probabilidad de validez, una precisión inicial (una probabilidad del resultado precalculado de ser válido en el momento en que se precalcula) de cualquiera de los resultados de búsqueda precalculados, etc., que se almacenan y actualizan permanentemente por el controlador 2 de recálculo u otra entidad acoplada al controlador 2 de recálculo. Un modelo probabilístico puede basarse en experiencia estadística del comportamiento de los resultados de búsqueda precalculados o puede formarse de manera conceptual en base a la experiencia de objeto. En general, un modelo probabilístico permite aproximar la validez de los resultados de búsqueda precalculados a lo largo del tiempo. Esta validez aproximada se denomina en lo sucesivo probabilidad de validez. En general, la probabilidad de validez de un resultado de búsqueda precalculado disminuye con el tiempo que pasa después del precálculo del resultado de búsqueda.

En la Fig. 2 se representan dos funciones a modo de ejemplo de la probabilidad de validez que disminuye con el tiempo. La función 10 representa un resultado de búsqueda precalculado que potencialmente permanece con una probabilidad más alta de ser válido a lo largo del tiempo que otro resultado de búsqueda precalculado asociado a la función 11. Por ejemplo, el resultado de búsqueda precalculado representado por la función 10 tiene una probabilidad del 70 % de ser todavía válido 35 horas después de su último recálculo, mientras que el otro resultado de búsqueda precalculado caracterizado por la función 11 solo es válido hasta aproximadamente el 50 % 35 horas después de su último recálculo. Las funciones 10 y 11 también pueden representar conjuntos completos de resultados de búsqueda precalculados (como, por ejemplo, partes como se describe más adelante) y luego indican proporciones de los conjuntos de resultados de búsqueda precalculados que probablemente son válidos en un tiempo pasado desde el último recálculo del conjunto.

Los modelos probabilísticos, sin embargo, no reflejan eventos inesperados o impredecibles que pueden disminuir sustancialmente la validez de los resultados de búsqueda precalculados, es decir, porciones significativas de los resultados de búsqueda precalculados pueden invalidarse tras la ocurrencia de un evento. Debido a sus características de ser inesperados o impredecibles, tales eventos generalmente no se incluyen en modelos probabilísticos. Suponiendo que el ejemplo de los resultados de búsqueda calculados previamente sean datos relacionados con viajes como, por ejemplo, recomendaciones de viaje con precios, ejemplos de tales eventos que tienen un impacto en la validez de los resultados de búsqueda calculados previamente son situaciones de la vida real como, por ejemplo, una feria comercial o un evento deportivo (que, p. ej., aumenta el precio de los vuelos en ciertas fechas y para ciertas ubicaciones) o eventos aleatorios como, por ejemplo, huelgas o desastres naturales (que pueden conducir a cancelaciones de vuelos), todos los cuales cambian las presunciones subyacentes a las causalidades del modelo probabilístico. Suponiendo que otro ejemplo de los resultados de búsqueda precalculados son sitios web de Internet prerecogidos, ejemplos de eventos que conducen a la invalidez de partes de los resultados de búsqueda precalculados son una campaña política que provoca un bloqueo de ciertos (tipos de) sitios web o un apagón técnico que provoca que servidores de Internet ubicados en una determinada área geográfica no tengan conexión durante un período más largo. Si los resultados de búsqueda calculados previamente son, p. ej., datos geográficos y relacionados con el tiempo como, por ejemplo, niveles de agua de río y mar o información sobre la contaminación del aire, una porción de los resultados de búsqueda calculados previamente puede invalidarse debido a un desastre natural como, por ejemplo, un tsunami o la erupción de un volcán.

Soluciones conocidas que consideran dichos eventos asíncronos en tiempo real que invalidan potencialmente ciertas partes de los resultados de búsqueda calculados previamente dependen de una señalización externa de dichos eventos, como se describe, p. ej., en el documento WO 2014/026753. Sin embargo, tal señalización externa puede no ser siempre posible o deseada, p. ej., debido a la ausencia de sistemas interconectados, interfaces técnicas adecuadas o impacto desconocido de los eventos en la validez de los resultados de búsqueda calculados previamente. Más allá de una señalización externa, el documento WO 2014/026753 también considera un reconocimiento implícito de eventos asíncronos en tiempo real mediante el empleo de un mecanismo de muestreo. Con este fin, ciertas partes representativas de los resultados de búsqueda precalculados (muestras) se vuelven a calcular intencionadamente de vez en cuando para determinar si una parte más grande de los resultados de búsqueda precalculados representados por una muestra tiene una validez real significativamente menor que la indicada por el modelo probabilístico. Sin embargo, determinar la validez real mediante tal proceso de muestreo requiere recursos de recálculo adicionales que se consumen a expensas del recálculo de los resultados de búsqueda precalculados más críticos según lo prescrito por la estrategia de recálculo, p. ej., los resultados de búsqueda precalculados con la probabilidad de validez más baja.

Ante el trasfondo de estos problemas tecnológicos de que los resultados de búsqueda precalculados se invalidan a lo largo del tiempo, las estrategias de recálculo sobre la base de modelos probabilísticos y la consideración y el reconocimiento de eventos en tiempo real que influyen en la validez de los resultados de búsqueda precalculados más allá de lo que refleja el modelo probabilístico, se propone en la presente memoria una manera eficiente (en términos de recursos informáticos) para reconocer implícitamente disminuciones de validez dentro de los resultados de

búsqueda precalculados, que están posiblemente causados por eventos asíncronos en tiempo real. No son necesarios ni la señalización externa ni el muestreo intensivo de recursos informáticos.

En resumen, a nivel general, la presente solución se basa en

- 5 - la determinación de un parámetro de modelo probabilístico en forma de una tendencia de validez que indica una tasa de disminución o tasa de retención de la probabilidad de validez de los resultados de búsqueda precalculados a lo largo del tiempo (p. ej., como se indica por las funciones a modo de ejemplo de la Fig. 2),
- la determinación de una tasa de validez instantánea que indica una validez instantánea de una primera parte de los resultados de búsqueda precalculados y un resultado de búsqueda precalculado dentro de esta primera parte, respectivamente,
- 10 - una comparación de la tasa de validez instantánea con la tendencia de validez y una determinación de si la diferencia entre la tasa de validez instantánea difiere o no de la tendencia de validez más de un grado dado,
- un ajuste de las probabilidades de validez de los resultados de búsqueda precalculados de una segunda parte diferente de, pero correlacionada con, la primera parte, si la diferencia entre la tasa de validez instantánea y la validez excede el grado dado, y
- 15 - un recálculo de los resultados de búsqueda precalculados según una estrategia de recálculo dada dependiente de las probabilidades de validez asociadas a los resultados de búsqueda precalculados. En consecuencia, las probabilidades de validez potencialmente ajustadas de los resultados de búsqueda precalculados en la segunda parte se tienen en cuenta cuando se decide qué resultados de búsqueda precalculados deben recalcularse.

20 Por lo tanto, la idea principal de este mecanismo es usar correlaciones de probabilidad de validez entre los resultados de búsqueda precalculados y transferir una discrepancia sustancial detectada entre la indicación de validez provista por el modelo probabilístico y una indicación de validez instantánea de una primera parte de resultados de búsqueda precalculados a una segunda parte de resultados de búsqueda precalculados correlacionados con la primera parte de resultados de búsqueda precalculados. De esta manera, por ejemplo, un proceso de muestreo para la segunda parte de resultados de búsqueda precalculados se vuelve obsoleto. En particular, es posible derivar la indicación de validez instantánea para la primera parte en respuesta a un recálculo "normal" de resultados de búsqueda precalculados en la primera parte, significando "normal" en este caso que este recálculo está en línea con la estrategia de recálculo empleada en oposición a un recálculo de muestreo artificial adicional o similar. Por lo tanto, la sobrecarga de recálculo puede reducirse o evitarse por completo.

25 Antes de pasar a la descripción detallada de los ejemplos de implementación, estas características se describen a continuación con más detalle en primer lugar a nivel funcional.

30 Como ya se ha descrito más arriba, la plataforma 4 de búsqueda almacena los resultados de búsqueda calculados previamente. Una representación de los resultados de búsqueda precalculados también se mantiene por el controlador 2 de recálculo con fines de control de recálculo. En particular, el controlador 2 de recálculo mantiene datos de control para emplear la estrategia de recálculo. Los datos de control permiten que el controlador 2 de recálculo determine probabilidades de validez de los resultados de búsqueda precalculados y, p. ej., inicie regularmente el recálculo de una porción de los resultados de búsqueda precalculados según la estrategia de recálculo, p. ej., resultados de búsqueda precalculados que tienen la probabilidad de validez más baja. Los parámetros para determinar las probabilidades de validez de los resultados de búsqueda calculados previamente son, por ejemplo, el tiempo del último recálculo y una función de disminución para la probabilidad de validez como se muestra a modo de ejemplo en la Fig. 2, por ejemplo $e^{-\lambda_i t_i}$, λ_i que denota una tendencia de validez de un resultado de búsqueda i precalculado particular modelado por un modelo probabilístico y t_i que denota un tiempo (número de unidades de tiempo) pasado desde el último recálculo del resultado de búsqueda i precalculado. Mediante el almacenamiento de estos dos parámetros λ_i y t_i para cualquier resultado de búsqueda precalculado (o conjuntos de resultados de búsqueda precalculados), el controlador 2 de recálculo puede calcular la probabilidad de validez del resultado de búsqueda i precalculado en cualquier momento dado. El término "tendencia de validez", como se usa en la presente memoria, se refiere a una tasa de disminución de la probabilidad de validez a lo largo del tiempo (en ejemplos también cubiertos en la presente memoria, también puede representar una tasa de mantenimiento de la probabilidad de validez a lo largo del tiempo). Además, se observa que el controlador 2 de recálculo puede almacenar y mantener permanentemente los parámetros λ_i y la marca de tiempo de recálculo de cada resultado de búsqueda precalculado, pero no los parámetros t_i y probabilidad de validez resultante de las funciones a modo de ejemplo descritas más arriba. Más bien, los dos últimos valores pueden calcularse dinámicamente bajo demanda, como $t_i = \text{tiempo actual} - \text{recálculo}$ y el valor de probabilidad de validez depende de λ_i y t_i como se da, p. ej., por las funciones descritas más arriba.

35 En general, los resultados de búsqueda precalculados se subdividen en partes, como se muestra en la Fig. 3 que indica partes D, D', E, F, G y H a modo de ejemplo. Los resultados de búsqueda precalculados dentro de una parte 40 tienen características de invalidación similares, es decir, tienen tendencias de validez λ_i idénticas o similares y, por lo tanto, funciones de cambio idénticas o similares para la probabilidad de validez. Por ejemplo, si los resultados de búsqueda precalculados son recomendaciones de viaje con precios, una parte de resultados de búsqueda 45

precalculados puede formarse por todas las recomendaciones de viaje con precios con una ubicación de origen y destino particular y que tienen, p. ej., fechas de partida dentro de un período dado, p. ej., un mes (como, por ejemplo, recomendaciones de viaje entre el par de ciudades Niza - Nueva York con partida en agosto de 2014). Si, por ejemplo, los resultados de búsqueda calculados previamente son registros de datos que indican parámetros ambientales como, por ejemplo, niveles de agua y contaminación del aire, puede formarse una parte mediante resultados de búsqueda calculados previamente de una región geográfica particular. El número de resultados de búsqueda precalculados dentro de una parte, así como el número de partes, dependen de la realización particular del sistema 1 de base de datos. Una parte puede incluir cualquier número de resultados de búsqueda precalculados.

La descomposición de los resultados de búsqueda calculados previamente en una parte puede reflejarse por el modelo de base de datos y la estructura de la plataforma 4 de búsqueda. Por ejemplo, cada grupo de resultados de búsqueda precalculados que forman una parte puede mantenerse en una tabla dedicada (u otra estructura de base de datos adecuada). Como alternativa, los resultados de búsqueda calculados previamente pueden incluir un parámetro que indica la parte a la que pertenecen. En este caso, cada registro de base de datos incluye un campo de parte. Como alternativa adicional, las partes pueden definirse mediante una estructura de datos separada como, por ejemplo, una tabla de partes que define qué resultado de búsqueda precalculado está asociado a qué parte. Otras formas de definir las partes como, por ejemplo, una definición lógica por reglas sin un reflejo en el modelo y estructura de la base de datos también se cubren en la presente memoria y no se excluyen por estos ejemplos.

Además, también puede existir una correlación a nivel de partes. Los resultados de búsqueda precalculados de una primera parte, p. ej., la parte D de la Fig. 3, pueden tener tendencias de validez λ_i y, por lo tanto, disminuir las funciones para la probabilidad de validez similar a los resultados de búsqueda precalculados en otra parte, p. ej., la parte D' de la Fig. 3 (indicado en la Fig. 3 por la flecha continua entre la parte D y la parte D'). Por ejemplo, la tendencia de validez promedio de todos los resultados de búsqueda precalculados en la primera parte D está cerca de la tendencia de validez promedio de todos los resultados de búsqueda precalculados en la segunda parte D' ($\lambda_D = 0,1$ y $\lambda_{D'} = 0,11$) p. ej. y, además, los resultados de búsqueda precalculados en la primera parte D y los resultados de búsqueda precalculados en la segunda parte D' se interrelacionan en términos de comportamiento de validez, es decir, si una parte sustancial de los resultados de búsqueda precalculados en la primera parte D es inválida, es probable que una parte sustancial de los resultados de búsqueda precalculados en la segunda parte D' también sea inválida. Por ejemplo, si los resultados de búsqueda precalculados son recomendaciones de viaje con precios, la parte D puede incluir todas las recomendaciones de viaje con precios precalculadas para el par de ciudades Niza - Nueva York con fechas de partida en agosto de 2014 y la parte D' puede incluir las recomendaciones de viaje con precios precalculadas para el mismo par de ciudades Niza - Nueva York con fechas de viaje en septiembre de 2014. Si, por ejemplo, los resultados de búsqueda calculados previamente son registros de datos que indican parámetros ambientales como, por ejemplo, niveles de agua y contaminación del aire, las partes D y D' pueden formarse mediante registros de datos relacionados con regiones geográficamente adyacentes como, por ejemplo, dos municipios vecinos.

Además, puede haber diferentes niveles de correlación entre partes. Por ejemplo, las partes D y D' de la Fig. 3 y de los ejemplos dados a continuación están fuertemente correlacionadas y, por tanto, los resultados de búsqueda precalculados en la primera parte D y los resultados de búsqueda precalculados en la segunda parte D' pueden tener características de validez casi idénticas. Otra parte, por ejemplo, la parte F de la Fig. 3, puede aun correlacionarse con la parte D, pero la correlación entre las partes D y F puede ser menos fuerte que la correlación entre las partes D y D'. Por ejemplo, si los resultados de búsqueda precalculados son recomendaciones de viaje con precios, la parte D puede incluir todas las recomendaciones de viaje con precios precalculadas para el par de ciudades Niza - Nueva York con fechas de viaje en agosto de 2014 y la parte F puede incluir las recomendaciones de viaje con precios precalculadas para el mismo par de ciudades Niza - Nueva York con fechas de viaje en octubre de 2014. Si, por ejemplo, los resultados de búsqueda calculados previamente son registros de datos que indican parámetros ambientales como, por ejemplo, niveles de agua y contaminación del aire, las partes D y F pueden formarse mediante registros de datos relacionados con regiones que todavía están geográficamente cercanas entre sí, pero no directamente adyacentes. Un ejemplo específico para determinar la correlación entre dos partes de resultados de búsqueda precalculados se provee más abajo.

En aras de la ilustración, la siguiente descripción se refiere a un resultado de búsqueda i precalculado individual, a modo de ejemplo, que se incluye en la parte D, como se muestra en la Fig. 3. Sin embargo, las siguientes declaraciones se aplican igualmente a cualquier resultado de búsqueda precalculado incluido en cualquier parte.

Determinación de la tendencia de validez

En un aspecto, el controlador 2 de recálculo mantiene un modelo probabilístico para los resultados de búsqueda precalculados. Para cualquier resultado de búsqueda i precalculado, el controlador 2 de recálculo determina una tendencia de validez λ_i que indica una tasa de disminución de la probabilidad de validez del resultado de búsqueda i precalculado. Por ejemplo, el resultado de búsqueda i precalculado puede tener una tasa de disminución de validez del 10 % por hora, lo cual significa que la probabilidad de que i sea válido disminuye en un 10 % cada hora ($\lambda_i = 0,1$). En el momento de su primer cálculo o recálculo, i es generalmente válido al 100 % (suponiendo una precisión inicial de 1). Después de una hora, i sigue siendo válido con una probabilidad del 90 %. Después de dos horas, la validez de i es del 81 % (=90 % disminuido en otro 10 %). Después de tres horas, la validez probable de i es del 72,9 %, y así sucesivamente. Por lo tanto, la tendencia de validez λ_i es una medida de cuánto tiempo permanece válido el resultado

de búsqueda i precalculado o cuán rápido se vuelve inválido el resultado de búsqueda i precalculado debido a cambios de los datos originales subyacentes.

5 La tendencia de validez λ_i modela la validez basada en la experiencia a largo plazo del desarrollo de validez del resultado de búsqueda i precalculado. El valor de tendencia de validez λ_i para un resultado de búsqueda i precalculado particular se determina, por lo tanto, a partir de al menos tres recálculos pasados de i. Cuanto mayor es el número de recálculos anteriores tenidos en cuenta para la determinación de λ_i más fiable es el valor de λ_i . El controlador de recálculo almacena la tendencia de validez λ_i como parámetro de datos de control para el resultado de búsqueda i precalculado.

10 Una dificultad para determinar la tendencia de validez λ_i es el hecho de que el recálculo del resultado de búsqueda i precalculado generalmente no se produce en intervalos de tiempo equidistantes. En particular, períodos variables pueden pasar entre dos recálculos posteriores de i, dependiendo de la estrategia de recálculo empleada.

15 La Fig. 4a provee un ejemplo que muestra una serie de recálculos sucesivos del resultado de búsqueda i precalculado. En este ejemplo, el primer recálculo se produce cinco horas después del cálculo previo de i. El recálculo muestra que el resultado de búsqueda i precalculado ha cambiado. Sin embargo, el recálculo no indica en qué momento específico durante este intervalo de tiempo de cinco horas entre el cálculo anterior y el primer recálculo se invalida realmente el resultado de búsqueda i precalculado. A continuación, el siguiente (segundo) recálculo de i tiene lugar dos horas después. El recálculo muestra que i era todavía válido (es decir, el recálculo de i era realmente innecesario). Posteriormente, el resultado de búsqueda i precalculado no se recalcula después de que haya pasado otro período de siete horas. Este tercer recálculo indica de nuevo que i ha cambiado, es decir, i se invalidó en algún momento de desconocido en la ventana de siete horas. El (cuarto) recálculo final de i se produce entonces cuatro horas después. De nuevo, el resultado de búsqueda i precalculado ha cambiado mientras tanto, en algún momento en el intervalo de tiempo de cuatro horas.

Como se ilustra mediante este ejemplo, por lo tanto, generalmente es incierto en qué momento exacto se invalidó el resultado de búsqueda i precalculado si un recálculo de i después de un cierto período indica que i ya no era válido.

25 Existen varias posibilidades para tratar este problema al determinar la tendencia de validez λ_i . Una opción es volver a calcular el resultado de búsqueda i precalculado regularmente en intervalos de tiempo equidistantes relativamente cortos durante un cierto período limitado (por ejemplo, una vez cada hora durante un período de cuatro semanas). De esta manera, λ_i puede determinarse de manera sistemática. Sin embargo, estos recálculos son artificiales y, por lo tanto, consumen una cantidad sustancial de recursos de recálculo adicionales.

30 Por lo tanto, de manera alternativa o adicional, se pueden emplear medidas estadísticas para derivar λ_i de los recálculos generalmente inequidistantes de i dentro de la estrategia de recálculo empleada. En algunas realizaciones, una estimación de probabilidad máxima de un valor esperado para λ_i se lleva a cabo para ello. La estimación de probabilidad máxima se basa en una serie de pares de recálculo de i. Cada par de recálculo incluye un tiempo desde el último recálculo de i y un indicador que indica si i ha cambiado o no desde el último recálculo de i. La Fig. 4b introduce una notación formal de tal serie de pares de recálculo según la cual el tiempo desde el último n-ésimo recálculo de i se denota como T_n y el indicador de cambio se indica como b_n . Con referencia de nuevo al ejemplo de la Fig. 4a, la serie de recálculo de la Fig. 4a puede especificarse como el conjunto de tuplas $\{T_1 = 300, b_1 = 0; T_2 = 120, b_2 = 1; T_3 = 420, b_3 = 0; T_4 = 240, b_4 = 0\}$, T_n dado en minutos y $b_n = 0$ que indica que el resultado de búsqueda i precalculado era inválido antes del recálculo y $b_n = 1$ que indica que el resultado de búsqueda precalculado aún era válido. En algunas realizaciones, el controlador 2 de recálculo mantiene tales datos reflejando los recálculos pasados como datos de control para cualquier resultado de búsqueda precalculado almacenado en la plataforma 4 de búsqueda y actualiza/extiende estos datos en el curso de cualquier recálculo.

45 La estimación de probabilidad máxima estima el valor de la tendencia de validez λ_i para un resultado de búsqueda i precalculado estableciendo la función de probabilidad del valor de tendencia de validez λ_i a estimar en línea con los datos históricos observados y determinando un valor máximo de esta función. El resultado es una "mejor estimación educada" del valor real de la tendencia de validez λ_i . En algunas de estas realizaciones, un valor de confianza para la tendencia de validez λ_i se estima por el método de máxima probabilidad. Este valor de confianza permite evaluar la desviación potencial del valor de λ_i estimada por el método de máxima probabilidad a partir del valor verdadero de λ_i . Esto se realiza, por ejemplo, utilizando la información de Fisher observada para evaluar la probabilidad de que el valor verdadero de λ_i se sitúa en un intervalo de confianza. Un ejemplo detallado de aplicación de la estimación de máxima probabilidad para estimar la tendencia de validez λ_i y obtener un factor de confianza se describe adicionalmente a continuación.

55 En otras realizaciones, la tendencia de validez λ_i se determina estableciendo una función de distribución empírica de la probabilidad de que el resultado de búsqueda i precalculado permanezca válido a lo largo del tiempo y luego revirtiendo la función de distribución empírica. Este enfoque estima empíricamente valores discretos para la probabilidad de validez en puntos particulares de tiempo después de un recálculo del resultado de búsqueda i precalculado. Para este fin, se pueden definir intervalos de tiempo que abarcan un punto de tiempo de recálculo. A continuación, las ventanas deslizantes de tamaño variable se desplazan a lo largo de los intervalos de tiempo y se

evalúa la probabilidad de que el resultado de búsqueda i precalculado no se invalide dentro del intervalo de tiempo para cada uno de los tamaños de ventana variables. Esto produce una secuencia de valores de probabilidad de validez individuales en puntos de tiempo particulares a partir de los cuales se deriva una función continua usando regresión. La opción de determinar la tendencia de validez λ_i estableciendo una función de distribución empírica se explica con más detalle más adelante con referencia a las Fig. 7 a 9.

Con el fin de lograr una implementación eficiente (en términos de tiempo de procesamiento y utilización de recursos de cálculo) de la determinación de la tendencia de validez λ_i , algunas realizaciones usan procesos MapReduce. MapReduce es un modelo de programación y una implementación correspondiente para procesar y generar grandes conjuntos de datos con un algoritmo distribuido paralelo, p. ej., mediante un sistema de base de datos. Un programa MapReduce incluye procedimientos Map y procedimientos Reduce. En primer lugar, un procedimiento Map() lleva a cabo el filtrado y la clasificación (como, por ejemplo, clasificación de recomendaciones de viaje con precios por pares origen-destino de partida, una cola para cada par origen-destino). En segundo lugar, se lleva a cabo un procedimiento Reduce() que realiza una operación de resumen (como, por ejemplo, contar el número de recomendaciones de viaje con precios en cada cola).

MapReduce también incluye ejecutar diversas tareas en paralelo, gestionar comunicaciones y transferencias de datos entre las diversas partes del sistema de base de datos de implementación, y facilita beneficiosamente la ejecución en paralelo, la escalabilidad, la redundancia y la tolerancia a fallas. Por lo tanto, el presente ejemplo se puede usar ventajosamente si la tendencia de validez λ_i (y/o la tasa de validez instantánea $\hat{\lambda}_i$) de un gran número de resultados de búsqueda precalculados (p. ej., cientos de miles o millones de recomendaciones de viajes con precios que implican datos históricos de recálculo correspondientes, p. ej., en el orden de los últimos 100 recálculos de cada recomendación de viaje con precios) que se van a determinar. Para implementar el presente ejemplo se puede usar cualquier biblioteca MapReduce disponible como, por ejemplo, transportada por Apache Hadoop. Un ejemplo particular de empleo del paradigma MapReduce para determinar la tendencia de validez λ_i mediante el establecimiento y la regresión de la función de distribución empírica se provee más abajo.

Se observa que la tendencia de validez λ_i no es necesariamente un número que sea específico para un resultado de búsqueda i individual precalculado. Más bien, la tendencia de validez λ_i puede estar describiendo la probabilidad de validez (después de un cierto tiempo transcurrido desde el último recálculo) de un conjunto completo de resultados de búsqueda precalculados como, por ejemplo, una parte de resultados de búsqueda precalculados, una subporción de una parte de resultados de búsqueda precalculados o varias partes de resultados de búsqueda precalculados. Por ejemplo, un valor de tendencia de validez λ puede ser un valor agregado para un conjunto de resultados de búsqueda precalculados como, por ejemplo, una tendencia de validez común λ_{Davg} para los resultados de búsqueda precalculados de la primera parte D. Para este fin, se consideran juntos varios resultados de búsqueda precalculados relacionados para obtener su valor de tendencia de validez común. Con referencia de nuevo al ejemplo de la Fig. 4b descrito más arriba, no solo se tiene en cuenta el historial de recálculo del resultado de búsqueda i precalculado $\{T1, b1; T2, b2; T3, b3; T4, b4\}$ (es preciso ver la Fig. 4a), sino que también se tiene en cuenta el historial de recálculo de un resultado de búsqueda precalculado i' relacionado con i : $\{T1', b1'; T2', b2'; T3', b3'\}$. Un valor de tendencia de validez común para ambos resultados de búsqueda precalculados i e i' se estima en base a $\{T1, b1; T2, b2; T3, b3; T4, b4; T1', b1'; T2', b2'; T3', b3'\}$. Esto puede generalizarse a más de dos resultados de búsqueda precalculados (siempre que estén relacionados como se describió más arriba con referencia a la Fig. 3) como, por ejemplo, todos los resultados de búsqueda precalculados dentro de una parte D. Para obtener números estadísticamente relevantes, generalmente hay una compensación entre valores de tendencia de validez específicos para uno (o pocos) resultados de búsqueda precalculados que requieren basar la estimación en un historial de recálculo más largo (para tener un número significativo de tuplas) y estimar valores de tendencia de validez comunes a un mayor número de resultados de búsqueda precalculados relacionados que solo requieren tener en cuenta un historial de recálculo más reciente (a expensas de significación/especificidad de los valores de tendencia de validez).

Por lo tanto, en algunas realizaciones, la tendencia de validez λ_i de un resultado de búsqueda i precalculado particular incluido en la primera parte D se deriva de la tendencia de validez común λ_{Davg} para los resultados de búsqueda precalculados de la primera parte D, p. ej., $\lambda_i = \lambda_{Davg}$. En estas realizaciones, las tendencias de validez se determinan por tanto solo a nivel de partes, pero no para resultados de búsqueda individuales.

Determinación de la tasa de validez instantánea

Mientras que la tendencia de validez λ_i indica una tasa de cambio a largo plazo del resultado de búsqueda i precalculado, la tasa de validez instantánea $\hat{\lambda}_i$ indica una tasa de validez reciente. A diferencia de la tendencia de validez λ_i que se determina sobre la base de datos históricos más extensos (incluidos al menos tres recálculos pasados del resultado de búsqueda i precalculado, pero generalmente un número significativamente mayor de recálculos como, por ejemplo, 100, 200, 500 o 1000 o más recálculos), la tasa de validez instantánea $\hat{\lambda}_i$ se calcula teniendo en cuenta exactamente dos recálculos solamente.

Como se ha explicado más arriba con referencia a la Fig. 4, el recálculo de un resultado de búsqueda i individual precalculado produce resultados binarios en términos de si el resultado de búsqueda precalculado ha sido válido o no en el momento del recálculo: el resultado de búsqueda i precalculado ha sido válido (el indicador b_i se establece, p. ej., en 1) o se descubre que el resultado de búsqueda precalculado es inválido y el valor recalculado de i difiere del valor anterior de i antes del recálculo (y el indicador b_i ese establece, p. ej., en 0).

Por lo tanto, para determinar un valor significativo para la tasa de validez instantánea $\hat{\lambda}_i$ del resultado de búsqueda i precalculado, se tienen en cuenta los recálculos de un número adicional de resultados de búsqueda precalculados que están relacionados con el resultado de búsqueda i precalculado. El término "relacionado" se refiere aquí de nuevo a una correlación entre los resultados de búsqueda precalculados en términos de sus características de probabilidad de validez, como se ha descrito más arriba en el curso de la introducción del concepto de partes (Figura 3). Por lo

tanto, la tasa de validez instantánea $\hat{\lambda}_i$ del resultado de búsqueda i precalculado se determina generalmente al nivel de la parte en donde se incluye el resultado de búsqueda i precalculado. Por ejemplo, todos los resultados de búsqueda precalculados en una parte particular D se consideran relacionados entre sí. Si, por ejemplo, todos los resultados de búsqueda calculados previamente en una parte particular D se vuelven a calcular juntos, los dos últimos recálculos de todos estos resultados de búsqueda calculados previamente (denominados recálculo "actual" y "anterior") se tienen en cuenta y se determina un valor común para la tasa de validez instantánea $\hat{\lambda}_i$ para todos estos resultados de búsqueda precalculados en la parte D . En otro ejemplo, solo una subporción (representativa, p. ej., distribuida uniformemente) de los resultados de búsqueda precalculados en la parte D se consideran para determinar la tasa de validez instantánea $\hat{\lambda}_i$, incluso si todos los resultados de búsqueda precalculados en la parte D se recalculan durante el recálculo actual. En otro ejemplo, solo una parte de los resultados de búsqueda precalculados en la parte D se recalculan mediante el recálculo actual y se considera o bien toda o bien una subporción de esta parte de resultados de búsqueda precalculados para determinar la tasa de validez instantánea $\hat{\lambda}_i$. En un ejemplo adicional, solo una parte de los resultados de búsqueda precalculados en la parte D se vuelven a calcular mediante el recálculo actual, pero también otros resultados de búsqueda precalculados que no se han recalculado mediante el recálculo actual, sino que solo se tienen en cuenta mediante un recálculo anterior para determinar la tasa de validez instantánea $\hat{\lambda}_i$. Los dos últimos recálculos de cualquier resultado de búsqueda precalculado relacionado con el resultado de búsqueda i precalculado se tienen en cuenta para determinar la tasa de validez instantánea $\hat{\lambda}_i$ independientemente de en qué momento se hayan producido los dos últimos recálculos de los resultados de búsqueda precalculados relacionados.

El número de resultados de búsqueda precalculados considerado para determinar la tasa de validez instantánea $\hat{\lambda}_i$ depende del número de resultados de búsqueda precalculados en la parte D en donde se incluye el resultado de búsqueda i precalculado. El número de resultados de búsqueda precalculados relacionados considerado para determinar la tasa de validez instantánea $\hat{\lambda}_i$ puede, por ejemplo, estar en el intervalo de 10, 50, 100, 1000, 10.000, 50.000 o 100.000 o más resultados de búsqueda precalculados. Cuanto mayor es el número de resultados de búsqueda precalculados considerados para determinar la tasa de validez instantánea $\hat{\lambda}_i$, mayor es la significación estadística de la tasa de validez instantánea $\hat{\lambda}_i$. Por lo tanto, el controlador 2 de recálculo puede emplear un umbral mínimo para el número de resultados de búsqueda precalculados que se tienen en cuenta para determinar la tasa de validez instantánea $\hat{\lambda}_i$.

De manera similar, como se ha ilustrado más arriba con referencia a la Fig. 4, los intervalos de recálculo sucesivos de un resultado de búsqueda individual precalculado generalmente no son equidistantes. Lo mismo es generalmente cierto también para el último intervalo de recálculo (es decir, el período entre los dos últimos recálculos) de múltiples resultados de búsqueda precalculados relacionados que se consideran para determinar la tasa de validez instantánea $\hat{\lambda}_i$. Esto se muestra mediante los ejemplos de las Fig. 5a y 5.

La Fig. 5a muestra un ejemplo con cuatro resultados de búsqueda precalculados, numerados de 1 a 4 (el número de resultados de búsqueda precalculados mostrados es pequeño den aras de la ilustración). Cada uno de estos resultados de búsqueda precalculados se recalcula según la estrategia de recálculo empleada en diferentes puntos de tiempo. Esto también conduce a diferentes longitudes de los intervalos de recálculo entre el primer recálculo y el segundo recálculo mostrados por la Fig. 5a. Por ejemplo, el intervalo de recálculo de los resultados de búsqueda precalculados 1 y 2 es de tres horas (aunque los puntos de recálculo de los resultados de búsqueda precalculados 1 y 2 difieren entre sí) ya que la estrategia de recálculo prescribía que ya era necesario un recálculo adicional después de tres horas. El intervalo de recálculo del resultado de búsqueda precalculado 3 es de cinco horas, el del resultado de búsqueda precalculado 4 es de siete horas. La Fig. 5b representa de nuevo una designación formal de los recálculos con T_i denotando la duración del último intervalo de recálculo del resultado de búsqueda i precalculado y b_i siendo el indicador que indica si se ha descubierto que el respectivo resultado de búsqueda i precalculado es válido o no en el segundo recálculo.

Debido a los últimos intervalos de recálculo generalmente diferentes e inequidistantes, la determinación de la tasa de validez instantánea $\hat{\lambda}_i$ se enfrenta a cuestiones similares a la determinación de la tendencia de validez como se describió más arriba porque generalmente es incierto en qué punto exacto del tiempo dentro del último intervalo de recálculo el respectivo resultado de búsqueda precalculado se ha vuelto inválido. Por estas razones, los mismos

mecanismos que se han descrito más arriba con respecto a la determinación de la tendencia de validez también pueden emplearse para determinar la tasa de validez instantánea $\hat{\lambda}_i$ (como, por ejemplo, la estimación de probabilidad máxima y el establecimiento de la función de distribución empírica). De manera similar a lo que se ha descrito más arriba para la determinación, también la determinación de la tasa de validez instantánea $\hat{\lambda}_i$ puede utilizar procesos MapReduce. Para evitar repeticiones, se remite a las explicaciones respectivas anteriores y a los ejemplos más específicos que se proveen a continuación.

Opcionalmente, la determinación de la tasa de validez instantánea $\hat{\lambda}_i$ para el resultado de búsqueda i precalculado se lleva a cabo en respuesta a un recálculo del resultado de búsqueda i precalculado. Más específicamente, la determinación de la tasa de validez instantánea $\hat{\lambda}_i$ para el resultado de búsqueda i precalculado se lleva a cabo en respuesta a un recálculo del resultado de búsqueda i precalculado que se inició mediante la estrategia de recálculo habitual empleada, es decir, el controlador 2 de recálculo determinó que era necesario un recálculo del resultado de búsqueda i precalculado en vista de la probabilidad de validez asociada a i . De esta manera, los recálculos artificiales como, por ejemplo, el proceso de muestreo descrito más arriba, son innecesarios y se utilizan recursos de recálculo de una manera eficiente.

Sin embargo, como se ha explicado anteriormente, la determinación de la tasa de validez instantánea $\hat{\lambda}_i$ para el resultado de búsqueda i precalculado también puede considerar los dos últimos recálculos de otros resultados de búsqueda precalculados relacionados con i aunque estos otros resultados de búsqueda precalculados no se recalcularon durante el recálculo actual, sino que su último recálculo tuvo lugar en un punto de tiempo anterior. Con referencia de nuevo a las Fig. 5a y 5b, el resultado de búsqueda precalculado 3 se recalcula mediante el recálculo actual y la tasa de validez instantánea $\hat{\lambda}_3$ para el resultado de búsqueda 3 precalculado se determina en consecuencia basándose en el valor de b_3 . Para este fin, también los dos últimos recálculos de los resultados de búsqueda precalculados 1, 2 y 4 y los valores correspondientes de b_1 , b_2 y b_4 se consideran para la determinación de $\hat{\lambda}_3$, dado que los resultados de búsqueda precalculados 1, 2 y 4 están relacionados con el resultado de búsqueda precalculado 3. Como se representa en la Fig. 5a, el último recálculo de los resultados de búsqueda precalculados 1, 2 y 4 se produjo en un punto de tiempo anterior al recálculo actual del resultado de búsqueda precalculado 3. Por otro lado, el propósito de la tasa de validez instantánea $\hat{\lambda}_i$ es indicar una tasa de validez reciente, más actualizada, similar a una instantánea, puede ser deseable incluir solo resultados de búsqueda precalculados relacionados en la determinación de la tasa de validez instantánea $\hat{\lambda}_i$ que experimentó su último recálculo dentro de un cierto marco temporal pasado. Con este fin, el controlador 2 de recálculo mantiene opcionalmente un umbral de tiempo de recálculo que prescribe una cantidad máxima de tiempo pasado desde el último recálculo. En tales realizaciones, solo los resultados de búsqueda precalculados relacionados con el resultado de búsqueda i precalculado se tienen en cuenta para la determinación de la tasa de validez instantánea $\hat{\lambda}_i$ que tenía el último recálculo dentro del umbral de tiempo de recálculo.

Opcionalmente, el umbral de número mínimo y el umbral de tiempo de recálculo pueden emplearse en combinación.

Comparación de la tendencia de validez y la tasa de validez instantánea y ajuste de las probabilidades de validez de los resultados de búsqueda precalculados en la segunda parte

Después de haber determinado los valores de la tendencia de validez λ_i y tasa de validez instantánea $\hat{\lambda}_i$ el controlador 2 de recálculo lleva a cabo una comparación entre ambos parámetros. La comparación puede ser, por ejemplo, una comparación real de ambos valores, pero también puede incluir una cierta edad (hipotética) del resultado de búsqueda precalculado. En otras palabras, la comparación también puede llevarse a cabo como una comparación de probabilidades de validez en un cierto momento después del recálculo (como, por ejemplo, la probabilidad de validez a las 10 horas después del recálculo), una probabilidad de validez basada en el valor determinado de la tendencia de validez λ_i y probabilidad de validez basada en la tasa de validez instantánea $\hat{\lambda}_i$ del valor determinado. Ambos enfoques son equivalentes desde un punto de vista matemático.

La comparación produce una diferencia entre la tendencia de validez λ_i y tasa de validez instantánea $\hat{\lambda}_i$ o entre probabilidad de validez basada en el valor determinado de la tendencia de validez λ_i y la probabilidad de validez basada en la tasa de validez instantánea $\hat{\lambda}_i$ del valor determinado respectivamente. La diferencia determinada se compara entonces con un umbral de diferencia dado (como, por ejemplo, una diferencia de probabilidad de validez del 10 %). Si la diferencia supera el umbral, se supone que se ha producido un evento perjudicial para la validez del resultado de búsqueda i precalculado que no se refleja en el modelo probabilístico. En este caso, el controlador de recálculo procede a ajustar las probabilidades de validez asociadas a otros resultados de búsqueda precalculados correlacionados con el resultado de búsqueda i precalculado porque la validez de estos otros resultados de búsqueda precalculados también puede verse afectada negativamente por el evento detectado.

Generalmente, la correlación considerada para el ajuste de las probabilidades de validez de los otros resultados de búsqueda precalculados se sitúa a nivel de partes. Como se ha descrito más arriba con referencia a la Fig. 3, existen diferentes niveles de correlaciones entre partes. Por ejemplo, la parte D que es la parte del resultado de búsqueda i precalculado puede correlacionarse fuertemente con la parte D'. La correlación entre la parte D y la parte F puede ser a nivel medio. Y solo puede existir una correlación menor entre D y las otras partes representadas en la Fig. 3. Se puede emplear un umbral de correlación para decidir si dos partes están correlacionadas entre sí. Si la correlación entre partes supera el umbral, se considera que estas dos partes están correlacionadas entre sí.

Dado que la diferencia detectada entre tendencia de validez λ_i y tasa de validez instantánea $\hat{\lambda}_i$ significa que la validez reciente del resultado de búsqueda i precalculado es menor que la probabilidad de validez a largo plazo dada por el modelo probabilístico y esto, a su vez, implica que la validez actual de otros resultados de búsqueda precalculados en la segunda parte correlacionada D' es probable que esté por debajo de la probabilidad de validez a largo plazo, las probabilidades de validez de los otros resultados de búsqueda precalculados en la segunda parte D' deben reducirse. En algunas realizaciones, la cantidad de disminución de las probabilidades de validez de los otros resultados de búsqueda precalculados en la segunda parte D' depende de la cantidad de la diferencia detectada entre la tendencia de validez λ_i y tasa de validez instantánea $\hat{\lambda}_i$, es decir, cuanto mayor es la diferencia entre tendencia de validez λ_i y tasa de validez instantánea $\hat{\lambda}_i$, mayor es la disminución de las probabilidades de validez de los otros resultados de búsqueda precalculados en la segunda parte D'.

En algunas realizaciones, el grado de correlación entre dos partes D y D' se determina usando una regresión lineal para modelar una relación entre:

- una diferencia entre una tendencia de validez a nivel de parte λ_D y una tasa de validez instantánea a nivel de parte $\hat{\lambda}_D$, λ_D indicando una tasa de cambio de la probabilidad de los resultados de búsqueda precalculados en la primera parte D (por ejemplo, las tendencias de validez promedio de todos los resultados de búsqueda precalculados en D), siendo $\hat{\lambda}_D$ una tasa de validez instantánea relacionada con los resultados de búsqueda precalculados en la primera parte D (por ejemplo, una tasa de validez instantánea determinada en base a los dos últimos recálculos de todos los resultados de búsqueda precalculados en D), y

- una diferencia entre una tendencia de validez a nivel de parte $\lambda_{D'}$ y una tasa de validez instantánea a nivel de parte $\hat{\lambda}_{D'}$, $\lambda_{D'}$ indicando una tasa de cambio de la probabilidad de validez de los resultados de búsqueda precalculados en la segunda parte D' (por ejemplo, las tendencias de validez promedio de todos los resultados de búsqueda precalculados en D'), $\hat{\lambda}_{D'}$ siendo una tasa de validez instantánea relacionada con los resultados de búsqueda precalculados en la segunda parte D' (por ejemplo, una tasa de validez instantánea determinada en base a los dos últimos recálculos de todos los resultados de búsqueda precalculados en D').

Más específicamente, la correlación entre las dos partes D y D' a modo de ejemplo se determina empleando la varianza de los valores pasados de la tasa de validez instantánea $\hat{\lambda}_D$ en el tiempo y la covarianza entre la tasa de validez instantánea $\hat{\lambda}_D$ y la tasa de validez instantánea $\hat{\lambda}_{D'}$. Estos dos indicadores estadísticos se definen de la siguiente manera:

$$\sigma_D^2 = \text{varianza} (\hat{\lambda}_D)$$

$$\sigma_{DD'} = \text{covarianza} (\hat{\lambda}_D, \hat{\lambda}_{D'})$$

Si la covarianza $\sigma_{DD'}$ entre las dos tasas de validez instantánea $\hat{\lambda}_D$ y $\hat{\lambda}_{D'}$ a nivel de parte supera un umbral dado (como, por ejemplo, $\frac{\sigma_{DD'}}{\sigma_D \sigma_{D'}} > 0.5$), una diferencia entre la tendencia de validez a nivel de parte λ_D y la tasa de validez instantánea a nivel de parte $\hat{\lambda}_D$ probablemente significa que existe una diferencia similar para la parte D' entre la tendencia de validez a nivel de parte $\lambda_{D'}$ y una tasa de validez instantánea a nivel de parte $\hat{\lambda}_{D'}$.

En algunas realizaciones, las tendencias de validez asociadas a los otros resultados de búsqueda precalculados incluidos en la segunda parte D' se ajustan en dependencia del grado de correlación entre la primera parte D y la segunda parte D'. Para ello, se determina el grado de correlación entre cada dos partes D y D' mantenidas por la plataforma 4 de búsqueda, por ejemplo, en base a la covarianza entre la tasa de validez instantánea $\hat{\lambda}_D$ y la tasa de validez instantánea $\hat{\lambda}_{D'}$ como se ha descrito más arriba.

En algunas realizaciones, la cantidad de disminución de las probabilidades de validez de los resultados de búsqueda precalculados en la segunda parte D' es proporcional al grado de correlación entre la primera parte D y la segunda parte D'.

En otras realizaciones, la cantidad de disminución de las probabilidades de validez de los resultados de búsqueda precalculados en la segunda parte D' depende de ambos, la cantidad de diferencia detectada entre la tendencia de validez λ_D y tasa de validez instantánea $\hat{\lambda}_D$ en la primera parte D y el grado de correlación entre la primera parte D y la segunda parte D'. Mediante el uso de la regresión lineal, la correlación entre las dos partes D y D' con respecto a la diferencia entre los valores de tendencia de validez y la tasa de validez instantánea de sus resultados de búsqueda precalculados puede modelarse mediante la siguiente relación:

$$\hat{\lambda}_{D'} - \lambda_{D'} \approx \frac{\sigma_{DD'}}{\sigma_D^2} (\hat{\lambda}_D - \lambda_D)$$

Mediante la introducción de un valor ϵ residual, esta relación puede volver a concebirse en la siguiente fórmula:

$$\hat{\lambda}_{D'} - \lambda_{D'} = \frac{\sigma_{DD'}}{\sigma_D^2} (\hat{\lambda}_D - \lambda_D) + \epsilon_{DD'}$$

El parámetro residual $\epsilon_{DD'}$ inducido por esta regresión puede modelarse, por ejemplo, estableciendo que $\epsilon_{DD'}$ está por debajo de un valor dado $\epsilon_{99\%}$ 99 % del tiempo). Si la diferencia modelada entre la tendencia de validez a nivel parte $\lambda_{D'}$ y la tasa de validez instantánea a nivel de parte $\hat{\lambda}_{D'}$ (el término a la izquierda de la ecuación) es menor que un valor residual relativamente frecuente (como $\epsilon_{99\%}$), la cantidad de la diferencia $\hat{\lambda}_D - \lambda_D$ y la correlación entre las dos partes D y D' se considera insignificante y no se lleva a cabo ningún ajuste de las probabilidades de validez de los resultados de búsqueda precalculados en la segunda parte D'. De lo contrario, las probabilidades de validez de los resultados de búsqueda precalculados en la segunda parte D' se ajustan según la diferencia determinada $\hat{\lambda}_D - \lambda_D$ y la correlación entre las dos partes D y D'. El controlador 2 de recálculo puede aplicar la última fórmula a cualquier segunda parte D' mantenida por la plataforma 4 de búsqueda.

El ajuste (es decir, la disminución) de las probabilidades de validez puede llevarse a cabo de diversas maneras. En algunas realizaciones, los valores de tendencia de validez λ_i de los resultados de búsqueda precalculados en la segunda parte D' (y la tendencia de validez $\lambda_{D'}$, respectivamente) se disminuyen. En otras realizaciones, la antigüedad de los resultados de búsqueda precalculados en la segunda parte D' (es decir, el tiempo desde el último recálculo) se aumenta artificialmente, es decir, la antigüedad $t_{D'}$ de D' está adaptada a $t_{D'}^{ajustado}$ de modo que la función de probabilidad de validez de D' dada por la tasa de validez instantánea se refleja en la función de probabilidad de validez ajustada basándose en la tendencia de validez, es decir, $e^{-\lambda_{D'} t_{D'}} = e^{-\lambda_{D'} t_{D'}^{ajustado}}$. El valor de edad adaptado viene dado entonces por $t_{D'}^{ajustado} = \frac{\lambda_{D'} t_{D'}}{\lambda_{D'}}$.

En otras realizaciones, los valores de probabilidad de validez (derivados de al menos los valores de tendencia de validez y la edad) se disminuyen en un valor absoluto o relativo.

Un ejemplo de la comparación de la tendencia de validez λ_D y la tasa de validez instantánea $\hat{\lambda}_D$, las partes correlacionadas y la disminución de probabilidad de validez de otras partes correlacionadas se dan adicionalmente a continuación con referencia a las Fig. 10 y 11.

Activación del recálculo

Finalmente, el controlador 2 de recálculo genera y emite órdenes de recálculo a la plataforma 3 de cálculo para recalcular una parte de los resultados de búsqueda precalculados basándose en las probabilidades de validez asociadas a los resultados de búsqueda precalculados.

El presente enfoque para reconocer una búsqueda precalculada potencialmente inválida y considerar este reconocimiento en la estrategia de recálculo se ha descrito ahora a un nivel funcional más general. El proceso resultante llevado a cabo por el controlador 2 de recálculo se visualiza mediante la Fig. 6. En 12, el controlador 2 de recálculo determina la tendencia de validez λ_i para el resultado de búsqueda i precalculado. A continuación, en 13, el controlador de recálculo determina la tasa de validez instantánea $\hat{\lambda}_i$ que indica la tasa de validez reciente del resultado de búsqueda i precalculado. Posteriormente, en 14, el controlador 2 de recálculo compara la tendencia de validez λ_i con la tasa de validez instantánea $\hat{\lambda}_i$. Si la tasa de validez instantánea $\hat{\lambda}_i$ es esencialmente mayor que la tendencia de validez λ_i , es decir, la diferencia entre ambas supera un umbral dado (indicado por ">" en la Fig. 6), el proceso avanza a la actividad 15 en donde el controlador 2 de recálculo adapta las probabilidades de validez de los resultados de búsqueda precalculados que se correlacionan con el resultado de búsqueda i precalculado. Más específicamente, se adaptan las probabilidades de validez para los resultados de búsqueda precalculados incluidos en las partes D' que se correlacionan con la parte D del resultado de búsqueda i precalculado. La adaptación tiene en cuenta la extensión de la diferencia entre la tendencia de validez λ_i y la tasa de validez instantánea $\hat{\lambda}_i$ determinada en 14. Además, la adaptación de 15 puede tener en cuenta el grado de correlación entre el resultado de búsqueda i precalculado (o la parte D en donde i está incluido) y los otros resultados de búsqueda precalculados cuyas probabilidades de validez se

van a adaptar (o la(s) parte(s) D' en donde estos otros resultados de búsqueda precalculados están incluidos).

A continuación, en 16, el controlador de recálculo inicia el recálculo de los resultados de búsqueda precalculados basándose en las probabilidades de validez asociadas a los resultados de búsqueda precalculados que dependen de la estrategia de recálculo empleada. Por ejemplo, se generan órdenes de recálculo con respecto a resultados de búsqueda precalculados que tienen una probabilidad de validez menor que otros resultados de búsqueda precalculados que tienen una probabilidad de validez mayor. Esto puede, por ejemplo, implementarse usando valores umbral de las probabilidades de validez: los resultados de búsqueda precalculados con una probabilidad de validez por debajo de dicho valor umbral necesitan recalcularse. Por consiguiente, el controlador de recálculo genera y envía las respectivas órdenes de recálculo. Los resultados de búsqueda precalculados con una probabilidad de validez por encima de dicho valor umbral se consideran como probablemente válidos y, en consecuencia, no necesitan recalcularse. Por consiguiente, no se emiten órdenes de recálculo con respecto a estos resultados de búsqueda precalculados. También se pueden emplear estrategias de recálculo más sofisticadas, p. ej., como se describe en la solicitud europea 14290040.6.

Con referencia de nuevo a la Fig. 6, si la actividad 14 ha mostrado que la tasa de validez instantánea $\hat{\lambda}_i$ no es sustancialmente inferior a la tendencia de validez λ_i , es decir, que no se excedía el umbral de diferencia dado, no se llevó a cabo la actividad 15. En este caso, el controlador 2 de recálculo inicia el recálculo en 16 con valores originales no adaptados de las probabilidades de validez. Si se llevó a cabo la actividad 15, el controlador 2 de recálculo inicia el recálculo basándose en estas probabilidades de validez que se han adaptado en la actividad 15. La adaptación de las probabilidades de validez llevada a cabo en 15 puede provocar el recálculo de estos resultados de búsqueda precalculados cuyas probabilidades de validez se adaptaron (o una porción de estos resultados de búsqueda precalculados con probabilidades de validez adaptadas) porque la estrategia de recálculo puede prescribir que el resultado de búsqueda precalculado con las probabilidades de validez más bajas se recalcularán con prioridad (por supuesto, la estrategia de recálculo también puede tener en cuenta otros factores como, por ejemplo, la popularidad de los resultados de búsqueda precalculados). De esta manera, la determinación de resultados de búsqueda precalculados potencialmente inválidos lograda por las actividades 12 a 15 da como resultado una validez generalmente mayor de los resultados de búsqueda precalculados almacenados por la plataforma 4 de búsqueda.

En respuesta a la recepción de los resultados de búsqueda recalculados que son calculados por la plataforma 3 de cálculo en respuesta a la actividad 16, el controlador 2 de recálculo puede entrar en el siguiente ciclo (como se indica en la Fig. 6 por las flechas 8 y 9) comenzando con la determinación de los valores de tendencia de validez λ_i de los resultados i recalculados o, si la actividad 12 se omite para los resultados de búsqueda i recalculados (ya que los valores λ_i de tendencia de validez ya se determinaron en un momento anterior y se consideró innecesaria una actualización), determinando los valores de tasa de validez instantánea $\hat{\lambda}_i$ de los resultados de búsqueda i recalculados.

Aunque el proceso de la Fig. 6 se ha descrito como un proceso secuencial de las actividades 12 a 16, estas actividades no tienen que llevarse a cabo de una manera estrictamente secuencial en la práctica. Más bien, por ejemplo, las actividades 12 y 13 pueden ejecutarse mediante diferentes módulos de control que estiman y actualizan la tendencia de validez y los valores de tasa de validez instantánea para todos los resultados de búsqueda precalculados periódicamente y en paralelo (es preciso ver también el ejemplo de la Fig. 12 descrito más adelante). Por lo tanto, por ejemplo, las actividades 13, 14 y 15 pueden ser llevadas a cabo por un módulo de control en el orden que se muestra por la Fig. 6, pero las actividades 12 y 16 pueden ser llevadas a cabo por otros módulos de control en paralelo e independientes de las actividades 13 a 15.

Estimación de la función de distribución empírica

Con referencia ahora a la descripción más específica de ejemplos particulares que implementan las funciones más generales descritas más arriba, las Fig. 7 a 9 se refieren a un ejemplo de implementación particular para determinar la tendencia de validez λ_i y/o la tasa de validez instantánea $\hat{\lambda}_i$ a saber, la derivación de la función de distribución acumulativa empírica del resultado de búsqueda i precalculado, denotada como $F_i(t)$ que especifica la probabilidad de que el resultado de búsqueda i precalculado no se vuelva inválido dentro del tiempo t . Es preciso observar que la siguiente descripción se aplica de nuevo a cualquier resultado de búsqueda i precalculado y, alternativamente, también a conjuntos de resultados de búsqueda precalculados (por ejemplo, para completar partes de resultados de búsqueda precalculados o porción de partes).

La Fig. 7 provee en primer lugar una visión general del flujo del proceso. Generalmente, la función de distribución empírica se deriva de estadísticas de recálculos previos del resultado de búsqueda i precalculado. Por lo tanto, como una primera actividad, los cambios pasados del resultado de búsqueda i precalculado se compilan en 17. La base para esta compilación de cambios pasados es, por ejemplo, los datos de control mantenidos por el controlador de recálculo como se describió más arriba con referencia a la Fig. 4 (es decir, el conjunto de tuplas de especificación de la serie de recálculos pasados del resultado de búsqueda i precalculado). Además, o como alternativa, la plataforma 4 de búsqueda y/o el controlador 2 de recálculo (u otra entidad como, por ejemplo, una base de datos histórica) pueden mantener versiones históricas del resultado de búsqueda i precalculado, p. ej., los valores pasados del resultado de búsqueda i precalculado y marcas de tiempo asociadas que indican en qué momento o intervalo de tiempo se

almacenó el valor histórico respectivo de i en la plataforma 4 de búsqueda. El objetivo de la actividad 17 es obtener una estructura de datos del desarrollo histórico de la búsqueda i precalculada.

La segunda actividad 18 es estimar empíricamente valores discretos para la probabilidad de validez en puntos de tiempo particulares después de un recálculo del resultado de búsqueda i precalculado. Esto se logra, por ejemplo, definiendo intervalos de tiempo que abarcan un punto de tiempo de recálculo en el que se supone que el resultado de búsqueda precalculado era válido. Estos intervalos de tiempo se denominan en lo sucesivo "períodos de estabilidad". Como se desconoce el punto temporal exacto de una invalidación de un resultado de búsqueda precalculado entre dos recálculos, se supone, como hipótesis, que la invalidez ocurrió en el medio entre dos recálculos posteriores que dieron como resultado diferentes valores del resultado de búsqueda precalculado. Por lo tanto, un período de estabilidad que abarca el n -ésimo recálculo se define como el período que comienza en el medio del intervalo de tiempo entre $n-1$ -ésimo recálculo y el n -ésimo recálculo y que finaliza en el medio del intervalo de tiempo entre el n -ésimo recálculo y el $n+1$ -ésimo recálculo. Esto se visualiza mediante la Fig. 8a que muestra cuatro recálculos del resultado de búsqueda i precalculado en los tiempos T_0, T_1, T_2, T_3 , y T_4 . El recálculo en T_0 resultó en el valor P_0 para el resultado de búsqueda i precalculado. Los recálculos en T_1 y T_2 resultaron en los valores P_1 y P_2 , respectivamente, para el resultado de búsqueda i precalculado. El valor P_3 del resultado de búsqueda i precalculado fue el resultado del recálculo en el tiempo T_3 (la función de etapa creciente mostrada por la Fig. 8a indica que el valor del resultado de búsqueda i precalculado ha aumentado hasta T_2 , es decir, $P_0 < P_1 < P_2$, pero $P_3 = P_2$). Finalmente, el recálculo en el tiempo T_4 resulta en el valor P_4 que es inferior a P_3 . Los puntos de tiempo medio anteriores a estos tiempos de recálculo T_n que producen un P_n diferente de P_{n-1} se indican mediante $(T_n + T_{n-1}) / 2$ y se indican como C_n . Por lo tanto, en este ejemplo, el período de estabilidad que abarca el recálculo en el tiempo T_1 (que produjo $P_1 \neq P_0$) se define como $C_1 - C_0$, en donde C_1 - se da por $(T_1 + T_2) / 2$ y C_0 - se da por $(T_0 + T_1) / 2$. Por lo tanto, como presunción para la derivación de $F_i(t)$, se supone que el resultado de búsqueda i precalculado era válido dentro del período de estabilidad $C_1 - C_0$ con el valor P_1 , es decir, el valor P_0 se invalidó en C_0 y el valor P_1 se invalidó en C_1 . De manera similar, el período de estabilidad que abarca el tiempo de recálculo T_2 (que produjo $P_2 \neq P_1$) se define como $C_2 - C_1$, en donde C_2 se da por $(T_4 + T_3) / 2$ porque solo el recálculo en T_4 (pero no el recálculo en el tiempo T_3) produjo un resultado diferente ($P_2 = P_3 \neq P_4$). Por lo tanto, como presunción para la derivación de $F_i(t)$, se supone que el resultado de búsqueda i precalculado era válido dentro del período de estabilidad $C_2 - C_1$ con el valor $P_2 = P_3$, es decir, el valor P_1 se invalidó en C_1 y el valor $P_2 = P_3$ se invalidó solo en C_2 .

Para un conjunto de períodos de estabilidad que se definen sobre los recálculos históricos de interés, la probabilidad de que el resultado de búsqueda i precalculado no se invalide dentro de k unidades de tiempo se define entonces como

$$P(\text{sin cambio} | kdt) = \frac{\sum_{0 \leq x < n} \max\{C_{x+1} - C_x - kdt, 0\}}{C_n - C_0}$$

en donde dt indica una base de tiempo (p. ej., minutos u horas, 10 minutos, 0,5 horas o 2 horas, etc.) y k indica un número de unidades de tiempo según la base de tiempo (p. ej., $k=3$ y siendo la base de tiempo 10 minutos, $3dt = 3 \times 10 = 30$ minutos).

El efecto de esta fórmula se visualiza mediante la Fig. 8b. En este ejemplo, un único período de estabilidad $C_1 - C_0$ se define con $C_0 = 20$ minutos y $C_1 = 70$ minutos, es decir, se supone que las invalidaciones del resultado de búsqueda i precalculado ocurren en el momento de $C_0 = 20$ minutos y $C_1 = 70$ minutos (no en el medio de este intervalo). La base de tiempo se elige para que sea de 10 minutos y $k = 2$ (es decir, 2×10 minutos = 20 minutos). Por lo tanto, kdt forma una ventana de tiempo de 20 minutos. Para determinar si un recálculo en kdt después de un (primer recálculo) produce o no un valor diferente del resultado de búsqueda i precalculado, esta ventana de tiempo se desliza entonces durante el período de estabilidad (indicado por las ventanas $2dt$ progresivas en la Fig. 8b). Durante las tres primeras posiciones, el resultado de búsqueda i precalculado permanece válido porque el resultado de búsqueda precalculado actualizado en C_0 se supone que solo cambia en C_1 (no antes). Sin embargo, durante las dos últimas posiciones de la ventana deslizante, se detecta una invalidez del resultado de búsqueda i precalculado porque la ventana $2dt$ alcanza o se superpone con el siguiente cambio asumido del resultado de búsqueda i precalculado en C_1 . Por lo tanto, la probabilidad de que el resultado de búsqueda i precalculado no se invalide en ningún período de 20 minutos en el período de estabilidad $C_1 - C_0$ es $3/5$, ya que en tres de las cinco posiciones de ventana deslizante no se detecta una invalidez de i . Con referencia de nuevo a la ecuación anterior, esto viene dado por $P(\text{sin cambio} | k=2) = (70-20-2 \cdot 10)/(70-20) = 3/5$. Es preciso observar que este es un ejemplo simplificado en aras de la explicación. En la práctica, generalmente existirán múltiples períodos de estabilidad (para cada par de recálculos que producen diferentes valores para el resultado de búsqueda precalculado). Por lo tanto, el mecanismo de ventana deslizante como se ha expuesto más arriba se llevará a cabo para cada período de estabilidad existente y se procesará mediante la ecuación anterior.

La estimación de los valores de probabilidad de validez discretos (actividad 17) incluye variar el número de unidades de base de tiempo k y calcular los valores de probabilidad de validez para cualquier k dada (p. ej., si la base de tiempo es de 10 minutos, el tamaño de ventana deslizante puede variarse comenzando desde 10 minutos, es decir, $k=1$, hasta 48 horas, es decir, $k=288$).

Dicho de otro modo, el objetivo de este algoritmo es determinar la probabilidad de que los dos recálculos posteriores del resultado de búsqueda i precalculado sean idénticos bajo la suposición de que el momento de invalidación del resultado de búsqueda i precalculado (C_0, C_1, \dots) se conoce y si el primer recálculo del resultado de búsqueda i precalculado se lleva a cabo en cualquier momento t_0 entre C_0 y C_1 y el segundo recálculo del resultado de búsqueda i precalculado se produce algún tiempo después (es decir, kdt posterior, en t_1).). Para esta estimación, diferentes t_0 posibles entre C_0 y C_1 se toman y para cada uno de estos primeros tiempos de recálculo se comprueba si el segundo recálculo en t_1 (es decir, $t_0 + kdt$) produce o no un resultado de búsqueda i precalculado diferente, es decir, si existía una invalidación en C_i entre t_0 y $t_0 + kdt$. Si, p. ej., para el 80 % de los tiempos ensayados (deslizantes) de t_0 , el segundo recálculo del resultado de búsqueda i precalculado en $t_1 = t_0 + kdt$ produce el mismo resultado de búsqueda i precalculado, se considera que la probabilidad de que el resultado de búsqueda i precalculado sea aún válido kdt después de su cálculo es del 80 %, es decir, $P(\text{válido después de } kdt) = 0,8$. A continuación, se lleva a cabo el mismo proceso con valores variados de k . Esto produce los valores discretos de probabilidad de validez como, por ejemplo, $P(\text{válido después de 1 hora}) = 0,990$, $P(\text{válido después de 2 horas}) = 0,981$, $P(\text{válido después de 3 horas}) = \dots$, y así sucesivamente.

Por lo tanto, de esta manera, el resultado de la actividad 17 es una serie de valores de probabilidad de validez individuales en puntos de tiempo particulares como se indica por la Fig. 9a.

Con referencia ahora de nuevo a la Fig. 7, la última actividad 18 del proceso de alto nivel de la estimación de función de distribución empírica se forma mediante una regresión exponencial de los valores de probabilidad de validez individuales con el fin de determinar la función de mejor ajuste correspondiente a los valores de probabilidad de validez discretos resultantes de la actividad 17. Esto se realiza mediante métodos de regresión comúnmente conocidos. La regresión da como resultado un valor para λ_i y $\hat{\lambda}_i$, respectivamente, como se indica en la Fig. 9b.

A continuación, se describe un ejemplo más específico de determinación de la tendencia de validez λ_i estableciendo $F_i(t)$. En este ejemplo, los resultados de búsqueda precalculados son datos relacionados con viajes, más particularmente recomendaciones de viajes con precios. La plataforma 4 de búsqueda precalcula y almacena los precios más bajos para cada viaje ofrecido. Sin embargo, el siguiente ejemplo no es específico para tales datos relacionados con viajes, sino que también puede transferirse a otros tipos de resultados de búsqueda precalculados.

Una característica de este ejemplo es que las actividades 17 y 18 son llevadas a cabo por dos trabajos MapReduce secuenciales. El primer trabajo MapReduce del presente ejemplo se refiere a la actividad 17 y lleva a cabo una reconstrucción del historial de precios de las recomendaciones de viaje con precios almacenadas por la plataforma 4 de búsqueda y gestionadas por el controlador 2 de recálculo. Cada recomendación de viaje con precios almacenada por la plataforma 4 de búsqueda tiene varios campos de datos como, por ejemplo, origen, destino, fecha de partida, fecha de regreso, ID de la oficina de viajes que ofrece la recomendación y precio. El controlador 2 de recálculo mantiene datos de control adicionales asociados a la recomendación de viaje con precios como, por ejemplo, la marca de tiempo del último recálculo. Además, el controlador 2 de recálculo o, alternativamente, otra entidad accesible por el controlador 2 de recálculo como, por ejemplo, una base de datos histórica, mantiene versiones históricas de las recomendaciones de viaje con precios que incluyen precios y marcas de tiempo anteriores de recálculos anteriores. La clave de base de datos usada para la reconstrucción del historial de precios de las recomendaciones históricas de viaje con precios es, por ejemplo, la combinación del ID de oficina de viajes de campos de datos, origen, destino, fecha de partida y fecha de regreso. El objetivo de la reconstrucción del historial de precios es, para cada versión histórica de cada recomendación de viaje, obtener el precio más barato y la marca de tiempo asociada a cada recálculo. Esto se realiza mediante los siguientes procedimientos de Map/Reduce:

- El procedimiento Map asocia a cada clave una tupla de marca de tiempo y precio de recálculo.
- El procedimiento Reduce asocia a una clave una lista de tuplas clasificadas de marca de tiempo y precio de recálculo.

Un resultado a modo de ejemplo para una recomendación de viaje con precios a modo de ejemplo es el siguiente:

La clave es PARAF08AA,NCE,LON,15/02/2012,20/02/2012, en donde PARAF08AA es el ID de oficina, NCE es el origen (código de aeropuerto de Niza), LON es el destino (código de aeropuerto para área metropolitana de Londres), 15/02/2012 es la fecha de partida y 20/02/2012 es la fecha de regreso. La recomendación de viaje definida por esta clave está asociada a una lista de marcas de tiempo de recálculo de tuplas y el precio resultante del recálculo ordenado por marcas de tiempo de recálculo como, por ejemplo, (15/01/2012,150), (16/01/2012,160), (17/01/2012, 160), etc., indicando la primera cadena como, por ejemplo, 15/01/2012, la marca de tiempo de recálculo, especificando el segundo número como, por ejemplo, 150, el precio en euros (es preciso observar que las marcas de tiempo en este ejemplo son de una granularidad de días en aras de la simplicidad, en la práctica, la resolución de las marcas de tiempo es mayor como, por ejemplo, minutos o segundos). Por lo tanto, en este ejemplo, se muestran tres cálculos en tres días consecutivos. El precio más barato disponible para la recomendación de viaje a modo de ejemplo fue de 150 euros en primer lugar y aumentó en 10 euros en algún momento entre los dos primeros cálculos.

Se lleva a cabo un segundo trabajo MapReduce para implementar la estimación de validez empírica (actividad 18), p. ej., empleando las técnicas de ventana deslizante como se ha descrito en detalle más arriba. Para ello, se calcula para todos los valores factibles de k (p. ej., unidades de tiempo entre 10 minutos y 48 horas) el número de posiciones de ventana de tiempo deslizantes de todas las posiciones de ventana de tiempo posibles que contienen un cambio de precio. Esto se realiza en el ID, origen y destino de la oficina a nivel de clave y, además, en un intervalo de tiempo entre el recálculo y la fecha de partida (en lo sucesivo denominado intervalo de avance). Por ejemplo, la clave PARAF08AA,NCE,LON,15,30 se utiliza para determinar el número de cambios de precio en un intervalo de avance entre 15 y 30 días, es decir, una ventana de tiempo de 15 a 30 días antes de la fecha de partida de la recomendación de viaje.

- 5
- 10 La motivación para considerar los intervalos de avance es evitar una explosión del número de claves y aumentar la significación estadística de la información de cambio de precio agregada a nivel de clave. Además, se pueden definir grupos de intervalos de avance y los cambios de precio determinados se agregan para cada grupo de intervalo de avance. Un agrupamiento a modo de ejemplo de intervalos de avance es {[0-3],[4-6], [7-13], [14-20], [21-30], [31-60], [61-90], [91-120], [121-180], [181, 270], [271-361]}, es decir, el primer grupo se da por la fecha de partida de la ventana de tiempo hasta tres días antes de la fecha de partida, el segundo grupo se da por la ventana de tiempo de cuatro a seis días antes de la fecha de partida, etc.
- 15

Un resultado a modo de ejemplo simplificado del segundo trabajo MapReduce (determinación de los valores de probabilidad de validez discretos para valores variables de k) es el siguiente:

- 20 - duración total de los períodos de estabilidad (es decir, suma de todos los períodos de estabilidad de todas las tuplas de todas las recomendaciones de viaje a precios que satisfacen la clave PARAF08AA,NCE,LON,15,30 es de 120 días),
- se emplean cuatro ventanas temporales de deslizamiento diferentes variando $k = 1, 2, 3, 4$,
- para cada valor de k , se calcula: $\sum_{0 \leq k < n} \max\{C_{i+k} - C_i - kdt, 0\}$
- 25 - esto produce las siguientes probabilidades de validez para $k=1$: 110/120 (es decir, en 110 de las posibles posiciones de ventana deslizante se ha dado que el precio de la recomendación de viaje con precio no se modifica en el segundo recálculo, es decir, el precio era estable para k unidad de tiempo en 110 de 120 casos), para $k=2$: 95/120, para $k=3$: 72/120 y para $k=4$: 60/120.

Estimación de probabilidad máxima

- 30 Alternativa al ejemplo de implementación de determinación de la tendencia de validez λ_i y/o la tasa de validez instantánea $\hat{\lambda}_i$ estableciendo la función de distribución empírica, el mecanismo de la estimación de máxima probabilidad puede emplearse para determinar la tendencia de validez y/o la tasa de validez instantánea. De nuevo, la siguiente descripción se aplica a cualquier resultado de búsqueda i precalculado y, alternativamente, también a conjuntos de resultados de búsqueda precalculados (por ejemplo, para completar partes de resultados de búsqueda precalculados o porción de partes). La notación de tiempo de recálculo T_n e indicadores b_n para indicar si el resultado de búsqueda i precalculado era inválido o no en el momento del recálculo como se introdujo más arriba también se usan aquí.
- 35

- 40 La estimación de máxima probabilidad es un método de estimación de los parámetros de un modelo probabilístico. Cuando se aplica a un conjunto de datos y se le da un modelo probabilístico, la estimación de máxima probabilidad provee estimaciones para los parámetros del modelo. En general, la estimación de probabilidad máxima selecciona el conjunto de valores de los parámetros del modelo que maximiza la función de probabilidad. Por lo tanto, este enfoque maximiza la correspondencia del modelo probabilístico seleccionado con los datos observados.

- 45 Para el presente problema de determinar la tendencia de validez λ_i y/o la tasa de validez instantánea $\hat{\lambda}_i$, la estimación de probabilidad máxima se puede aplicar de la siguiente manera: la estimación se basa en una muestra de recálculos históricos del resultado de búsqueda i precalculado (o un conjunto de resultados de búsqueda precalculados como, por ejemplo, la parte D) que son los datos observados. El modelo probabilístico subyacente se define en que para un resultado de búsqueda i precalculado que tiene una tendencia de validez λ_i , la probabilidad de $b_n = 1$ (es decir, i sigue siendo válido) después del tiempo T_n es $e^{-\lambda_i T_n}$, y la probabilidad de $b_n = 0$ (es decir, i no es válido) es $1 - e^{-\lambda_i T_n}$. Por tanto, b_n sigue una ley clásica de Bernoulli:

$$b_n \sim \text{Be}(e^{-\lambda_i T_n})$$

50

Una muestra histórica de recálculos del resultado de búsqueda i precalculado puede definirse como $b = (b_1, b_2, b_3, \dots, b_n)$. La función de probabilidad L de esta muestra es la probabilidad de que esta muestra ocurra realmente para un parámetro dado λ_i :

$$L(\lambda_i) = P(b|\lambda_i) = \prod_{b_n=1} e^{-\lambda_i T_n} \prod_{b_n=0} (1 - e^{-\lambda_i T_n})$$

(es decir, sin cambios) (es decir, con cambios)

5 Si una tendencia de validez λ_i común y/o la tasa de validez instantánea $\hat{\lambda}_i$ para varios resultados de búsqueda precalculados (p. ej., todos los resultados de búsqueda precalculados dentro de una parte D)

Además, opcionalmente, se usa un factor de confianza para evaluar la desviación potencial de λ_i estimada por el método de máxima probabilidad a partir del valor verdadero de λ_i (por razones de claridad, el valor estimado de λ_i se denota como $\bar{\lambda}$ en lo sucesivo). Esto se realiza, por ejemplo, evaluando la probabilidad de que el valor verdadero de

10 λ_i está situado en el intervalo de confianza $[\bar{\lambda} - \delta, \bar{\lambda} + \delta]$. Para ello, por ejemplo, la información \bar{I}_F de Fisher observada se utiliza, lo cual se aplica a la presente situación, definida como

$$\bar{I}_F = \sum_{b_n=0} T_n^2 \frac{e^{-\bar{\lambda} T_n}}{(1 - e^{-\bar{\lambda} T_n})^2}$$

Con un tamaño creciente de la muestra, el valor verdadero de λ_i dada por la muestra observada $b, \lambda_i|b$, converge a una distribución normal con el valor $\bar{\lambda}$ esperado y varianza $\sigma^2 = 1/\bar{I}_F$, es decir

15
$$\lambda_i|b \rightarrow \mathcal{N}\left(\bar{\lambda}, \frac{1}{\bar{I}_F}\right)$$

Esto significa que, dada la muestra observada, la probabilidad de tener el parámetro verdadero λ_i en el intervalo $[\bar{\lambda} - \delta, \bar{\lambda} + \delta]$ puede evaluarse mediante:

$$P(\bar{\lambda} - \delta < \lambda_i < \bar{\lambda} + \delta) = P\left(\frac{-\delta}{\sigma} < \frac{\lambda_i - \bar{\lambda}}{\sigma} < \frac{\delta}{\sigma}\right) \approx 2\Phi\left(\frac{\delta}{\sigma}\right) - 1$$

en donde Φ es la función de distribución gaussiana acumulativa.

20 Como resultado, esto produce un factor de confianza de la estimación de probabilidad máxima que es:

$$\bar{I}_F = \sum_{b_n=0} T_n^2 \frac{e^{-T_n}}{(1 - e^{-\bar{\lambda} T_n})^2}$$

Este factor de confianza se puede usar para evaluar la probabilidad de λ_i que está fuera del intervalo dado por δ :

$$P(\lambda_i \text{ está fuera del intervalo } \delta) \approx 2\left(1 - \Phi\left(\delta\sqrt{\bar{I}_F}\right)\right)$$

25 Una aplicación particular de este factor de confianza es, por ejemplo, evaluar el tamaño de muestra adecuado para obtener una estimación segura por la estimación de máxima probabilidad. Por tanto, para

$$\bar{I}_F = \sum_{b_n=0} T_n^2 \frac{e^{-T_n}}{(1 - e^{-\bar{\lambda} T_n})^2}$$

Este factor de confianza se puede usar para evaluar la probabilidad de λ_i que está fuera del intervalo dado por δ :

$$P(\lambda_i \text{ está fuera del intervalo } \delta) \approx 2\left(1 - \Phi\left(\delta\sqrt{\bar{I}_F}\right)\right)$$

Una aplicación particular de este factor de confianza es, por ejemplo, evaluar el tamaño de muestra adecuado para obtener una estimación segura por la estimación de máxima probabilidad. Así, por ejemplo, en algunas realizaciones, el tamaño de la muestra se aumenta progresivamente hasta que la probabilidad de cometer un error del 20 % (es decir, $\delta = 0,2 \hat{\lambda}$) sea inferior al 10 %.

5 La Fig. 10 muestra un ejemplo de resultados de búsqueda precalculados almacenados por la plataforma 4 de búsqueda y los respectivos valores a modo de ejemplo para las probabilidades de validez en base a la tendencia de validez a largo plazo λ según el modelo probabilístico (en la Fig. 10 denotado como "Modelo") y la probabilidad de
 10 validez dada por la tasa de validez instantánea $\hat{\lambda}_i$ (en la Fig. 10 denominada "Instante"). Sin limitación, el ejemplo de la Fig. 10 se refiere de nuevo a resultados de búsqueda calculados previamente que son datos relacionados con viajes, concretamente recomendaciones de viaje con precios que indican inicio y origen, fecha de partida y regreso, el precio
 15 calculado previamente y otra información de viaje relevante como, por ejemplo, clase de reserva, etc. En el ejemplo de la Fig. 10, las recomendaciones de viaje con precios se agrupan en partes de la clave origen - destino - mes de fecha de partida, las partes mostradas en forma de una matriz (columnas designadas con una letra mayúscula, filas con un número, es decir, por ejemplo, las recomendaciones de viaje con precios con origen = París, destino = Nueva York, mes de partida = agosto se agrupan en la parte B1). En aras de la simplicidad, solo se muestra un número limitado de partes. En realidad, en general, puede estar presente un número significativamente mayor de partes.

La Fig. 10 ilustra una situación en la cual la tendencia de validez λ y la tasa de validez instantánea $\hat{\lambda}_i$ se han determinado para una porción de las partes (es decir, para las partes B1, E1, D2, B3 y C3) según las metodologías expuestas más arriba. Los valores para las probabilidades de validez basándose en la tendencia de validez λ indicada
 20 por la Fig. 10 son valores promedio para todas las recomendaciones de viaje con precios precalculadas en la parte respectiva. Así, por ejemplo, se ha determinado un valor de 0,87 en base a la tendencia de validez λ_{B1} de la parte B1, que significa que el valor promedio de las probabilidades de validez se basa en las tendencias de validez λ_i de todas las recomendaciones de viaje con precios precalculadas en la parte B1 es 0,87, mientras que los valores individuales de probabilidades de validez basándose en los valores individuales de λ_i asociadas a cada una de las
 25 recomendaciones de viaje con precios precalculadas en la parte B1 pueden ser menores o mayores. De manera similar, los valores de las probabilidades de validez basándose en la tasa de validez instantánea $\hat{\lambda}_i$ mostrada por la Fig. 10 son valores promedio. Los valores de probabilidades de validez de la tasa de validez instantánea $\hat{\lambda}_i$ resultan de los dos últimos recálculos de todas las recomendaciones de viaje con precios precalculadas de una parte o al menos de un subconjunto mínimo (representativo) dado de las recomendaciones de viaje con precios precalculadas
 30 de una parte. Por ejemplo, el valor de probabilidad de validez basado en la tasa de validez instantánea $\hat{\lambda}_{B1}$ para la parte B1 (0,85) se ha determinado recientemente en respuesta al último recálculo llevado a cabo por la plataforma 3 de cálculo que incluía todas las recomendaciones de viaje con precios precalculadas de la parte B1. Si, por ejemplo, la parte B1 incluye 1.000 recomendaciones de viaje con precios, el último recálculo de todas las 1.000 recomendaciones de viaje con precios precalculadas de la parte B1 indicará que aprox. 850 de las 1.000
 35 recomendaciones de viaje con precios han sido válidas en el momento del recálculo (es decir, el recálculo de estas 850 recomendaciones de viaje con precios no cambió el valor del precio), mientras que aprox. 150 de las 1.000 recomendaciones de viaje con precios eran inválidas (es decir, el recálculo de estas 150 recomendaciones de viaje con precios produjo un precio diferente del que estas 150 recomendaciones de viaje con precios tenían antes del recálculo). Por supuesto, enfoques más sofisticados para determinar la tasa de validez instantánea $\hat{\lambda}_i$ teniendo en cuenta los intervalos de recálculo generalmente inequidistantes de los resultados de búsqueda precalculados como se describió más arriba (la estimación de probabilidad máxima o el establecimiento de función de distribución empírica) pueden aplicarse para la determinación real de $\hat{\lambda}_i$. Por lo tanto, en el nivel individual de cada una de las
 40 recomendaciones de viaje con precios precalculadas en la parte B1, la probabilidad de validez dada por $\hat{\lambda}_i$ también tiene un valor de 0,85 (es preciso observar que esto no implica necesariamente que estos valores de $\hat{\lambda}_i$ se mantengan en realidad como datos de control por la plataforma 2 de recálculo, más bien, en el ejemplo de la Fig. 10, es suficiente que el valor a nivel de parte promedio de $\hat{\lambda}_i$ se almacene por el controlador 2 de recálculo). Alternativamente, el último ciclo de recálculo que incluye recomendaciones de viaje con precios precalculadas de la parte B1 solo afectó a 500 de las 1.000 recomendaciones de viaje con precios precalculadas de la parte B1 e indicó que 425 de estas 500 recomendaciones de viaje con precios precalculadas eran todavía válidas y 75 de estas 500 recomendaciones de viaje con precios precalculadas eran inválidas.
 50

X3 (todas las partes en la fila 3) están virtualmente no correlacionadas con las otras partes X1 y X2, p. ej., las partes X3 se refieren a viajes domésticos estadounidenses con origen = Boston y destino = Miami, mientras que las partes X1 y X2 se refieren a viajes que se originan en Europa.

En el ejemplo de la Fig. 10, los valores para la tendencia de validez λ de las partes B1, E1, B3 y C3 ya se han
 55 determinado en un punto de tiempo anterior. Los valores de la tasa de validez instantánea $\hat{\lambda}_i$ de las partes B1, E1, B3 y C3 también se han determinado anteriormente, p. ej., en los últimos recálculos respectivos de recomendaciones de viaje con precios precalculadas incluidas en las partes B1, E1, B3 y C3 (más particularmente, en los últimos recálculos respectivos que recalculan o bien todas las recomendaciones de viaje con precios precalculadas en las partes respectivas o al menos un subconjunto dado de las recomendaciones de viaje con precios precalculadas en las partes

respectivas). El último ciclo de recálculo activado por el controlador 2 de recálculo y ejecutado por la plataforma 3 de cálculo está ahora relacionado con las recomendaciones de viaje con precios precalculadas de la parte D2 (ya sea todas las recomendaciones de viaje con precios precalculadas de la parte D2 o un subconjunto mínimo (representativo) dado - por supuesto, el último ciclo de recálculo también puede haber recalculado otras recomendaciones de viaje con precios precalculadas de otras partes). En respuesta al recálculo de las recomendaciones de viaje con precios precalculadas de la parte D2, un valor actualizado para la tendencia de validez λ_{D2} se determina (actividad 12 de la Fig. 6 - es preciso observar que la determinación de la tendencia de validez λ_{D2} también se puede omitir aquí y se puede haber llevado a cabo en un punto anterior del tiempo, p. ej., en el curso de la inicialización o entrenamiento del modelo probabilístico - ya que la tendencia de validez es una tendencia a largo plazo, generalmente es suficiente determinar la tendencia de validez a intervalos significativamente más largos que los recálculos individuales de los resultados de búsqueda precalculados), lo cual aquí produce una probabilidad de validez de 0,85. Además, un valor de la tasa de validez instantánea $\tilde{\lambda}_{D2}$ se deriva del último recálculo (actividad 13 de la Fig. 6), que da como resultado solo una probabilidad de validez de 0,62. Por lo tanto, el último recálculo de, p. ej., 1.000 recomendaciones de viaje con precios precalculadas de la parte D2 indicó que solo aproximadamente 620 recomendaciones de viaje con precios precalculadas aún han sido válidas desde el recálculo anterior, pero se descubrió que aproximadamente 380 recomendaciones de viaje con precios precalculadas han sido invalidadas desde el recálculo anterior.

Después de haber determinado el valor actual de la tasa de validez instantánea $\tilde{\lambda}_{D2}$, el controlador 2 de recálculo lleva a cabo una comparación entre la tasa de validez instantánea $\tilde{\lambda}_{D2}$ y el valor de la tendencia de validez λ_{D2} (actividad 14 de la Fig. 6). Como se indica en la Fig. 10, esta comparación también se puede llevar a cabo en el nivel de los valores de probabilidad de validez resultantes de la probabilidad de validez de 0,85. Además, un valor de la tasa de validez instantánea $\tilde{\lambda}_{D2}$ se deriva del último recálculo (actividad 13 de la Fig. 6), que da como resultado solo una probabilidad de validez de 0,62. Por lo tanto, el último recálculo de, p. ej., 1.000 recomendaciones de viaje con precios precalculadas de la parte D2 indicó que solo aproximadamente 620 recomendaciones de viaje con precios precalculadas aún han sido válidas desde el recálculo anterior, pero se descubrió que aproximadamente 380 recomendaciones de viaje con precios precalculadas se han invalidado desde el recálculo anterior.

Después de haber determinado el valor actual de la tasa de validez instantánea $\tilde{\lambda}_{D2}$, el controlador 2 de recálculo lleva a cabo una comparación entre la tasa de validez instantánea $\tilde{\lambda}_{D2}$ y el valor de la tendencia de validez λ_{D2} (actividad 14 de la Fig. 6). Como se indica en la Fig. 10, esta comparación también se puede llevar a cabo a nivel de los valores de probabilidad de validez resultantes de los diferentes valores de $\tilde{\lambda}_{D2}$ y λ_{D2} . El controlador 2 de recálculo determina que se supera el umbral dado para la diferencia entre la tasa de validez instantánea $\tilde{\lambda}_{D2}$ y la tendencia de validez λ_{D2} (p. ej., una diferencia de probabilidad de validez de 0,1, 0,15 o 0,2). Por lo tanto, el controlador 2 de recálculo inicia la adaptación de las probabilidades de validez de las recomendaciones de viaje con precios precalculadas de las partes correlacionadas con la parte D2 (actividad 15 de la Fig. 6). Por ejemplo, como la parte E1 se correlaciona con la parte D2 a nivel del medio, el controlador 2 de recálculo adapta la probabilidad de validez de la parte E1 reduciendo el valor de la probabilidad de validez en un 12 % de 0,95 a, p. ej., 0,83. También la parte B1 se correlaciona con la parte D2, aunque esta correlación puede ser menos fuerte. Por lo tanto, el controlador 2 de recálculo también adapta la probabilidad de validez, p. ej., en aprox. 9 % de 0,87 a 0,79. La disminución de la probabilidad de validez puede lograrse, p. ej., adaptando el valor de las tendencias de validez (λ_{E1} , λ_{B1}), modificando la marca de tiempo del último recálculo de las recomendaciones de viaje con precios precalculadas respectivas o almacenando un valor de reducción absoluto o relativo como datos de control adicionales (p. ej., en el ejemplo de la parte E1, un valor de reducción de probabilidad de validez fijo y absoluto de 0,12).

Más adelante, el controlador 2 de recálculo toma la siguiente decisión de recálculo con el fin de seleccionar las recomendaciones de viaje con precios precalculadas que se deben volver a calcular a continuación según la estrategia de recálculo empleada (actividad 16 de la Fig. 6). Esta decisión se basa en las probabilidades de validez disminuidas, es decir, por ejemplo, en los valores disminuidos de λ_{E1} y λ_{B1} , lo cual puede provocar un recálculo más temprano de las recomendaciones de viaje con precios precalculadas de las partes E1 y B1 que sin la adaptación realizada por el controlador 2 de recálculo. Después de que se hayan recalculado las recomendaciones de viaje con precios precalculadas en la parte E1 y la parte B1, respectivamente, los valores para λ_{E1} y λ_{B1} opcionalmente se determinan de nuevo (p. ej., empleando los mecanismos de determinación de la función de distribución empírica o la estimación de máxima probabilidad como se ha descrito en detalle más arriba - actividad 12 de la Fig. 6) o, alternativamente, se restablecen a sus valores previos (dado que se puede esperar que la validez de las recomendaciones de viaje con precios precalculadas en la parte E1 y la parte B1 disminuya como normal después del recálculo). Además, también los valores para $\tilde{\lambda}_{E1}$ y $\tilde{\lambda}_{B1}$ se determinan de nuevo (actividad 13 de la Fig. 6) lo cual desencadena de nuevo la comparación de $\tilde{\lambda}_{E1}$ y $\tilde{\lambda}_{B1}$ con λ_{E1} y λ_{B1} , respectivamente (actividad 14 de la Fig. 6), y potencialmente una adaptación de las probabilidades de validez de las partes correlacionadas (actividad 15 de la Fig. 6).

La Fig. 11 visualiza la adaptación de la probabilidad de validez asociada a un resultado de búsqueda *i* precalculado (el término resultado de búsqueda *i* precalculado en lo sucesivo abarca tanto un resultado de búsqueda precalculado individual como un conjunto de resultados de búsqueda precalculados como, por ejemplo, una parte) como se describió más arriba junto con la actividad 15 de la Fig. 6. En el ejemplo de la Fig. 11, la línea 20 recta gruesa representa una probabilidad de validez que disminuye a una tasa de 0,01 por hora (es decir, el resultado de búsqueda

precalculado representado por la línea 20 permanece válido durante una hora con una probabilidad del 99 % o el 99 % del conjunto de resultados de búsqueda precalculados representado por la línea 20 permanece válido durante una hora). Por lo tanto, la línea 20 visualiza la probabilidad de validez del resultado de búsqueda i precalculado aproximado/modelado por la función $e^{-\lambda_i} t_i \approx e^{-0,01 t_i} (\lambda_i \approx 0,01$ que es una tasa de disminución de validez de 1 % por hora). Por lo tanto, 18 horas después del último recálculo del resultado de búsqueda i precalculado, la probabilidad de validez del resultado de búsqueda precalculado se modela como $e^{-0,01 \cdot 18} \approx 0,835$. En este momento, el controlador 2 de recálculo determina que la diferencia entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i de un resultado de búsqueda precalculado j que está correlacionado con el resultado de búsqueda i precalculado supera significativamente el umbral de diferencia dado. Como consecuencia, la probabilidad de validez del resultado de búsqueda i precalculado se reduce dependiendo de la cantidad de diferencia determinada entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i y, opcionalmente, dependiendo del nivel de correlación entre i y j , en el ejemplo de la figura de aprox. 0,835 por 30 % ($\approx 0,25$) a aproximadamente 0,585. La función 21 resultante se aproxima entonces a la nueva probabilidad de validez reducida del resultado de búsqueda i precalculado con la función $e^{-\lambda_i t_i} \dots - 0,25$.

La Fig. 12 ilustra la arquitectura interna de un controlador 2 de recálculo a modo de ejemplo que implementa las metodologías descritas más arriba. Según el ejemplo de la Fig. 12, el controlador 2 de recálculo incluye los siguientes componentes:

- El almacenamiento 26 de resultados de búsqueda precalculados es un repositorio de datos distribuidos que contiene los resultados de búsqueda precalculados históricos calculados por la plataforma 3 de cálculo durante un período dado como, por ejemplo, varios meses pasados. Los resultados de búsqueda calculados previamente son insertados en el almacenamiento 26 por el gestor 25 de repositorio que recibe los resultados de búsqueda calculados de nuevo de la plataforma 3 de cálculo.

- Evaluador 27 de tendencia de validez: este componente analiza las diferencias entre los recálculos sucesivos de resultados de búsqueda precalculados y genera los valores de tendencia de validez λ_i . Los valores de tendencia de validez λ_i (así como otros datos de control asociados a los resultados de búsqueda precalculados) se almacenan en la representación 30 de datos interna a través del gestor 31 de entrada.

- Evaluador 28 de tasa de validez instantánea: este componente es activado por el gestor 25 de repositorio cada vez que el gestor 25 de repositorio recibe un conjunto de resultados de búsqueda recalculados de la plataforma 3 de cálculo e inserta el conjunto de resultados de búsqueda recalculados en la representación 30 de datos interna. El evaluador 28 de tasa de validez instantánea determina los valores de tasa de validez instantánea $\hat{\lambda}_i$ comparando las dos últimas versiones de cada resultado de búsqueda precalculado. El evaluador 28 de tasa de validez instantánea almacena valores de tasa de validez instantánea $\hat{\lambda}_i$ asociados a los resultados de búsqueda precalculados en la representación 30 de datos interna a través del gestor 31 de entrada.

- Evaluador 29 de correlación: este componente determina la correlación entre partes de resultados de búsqueda precalculados. También puede ser responsable de gestionar las partes de resultados de búsqueda precalculados, es decir, subdividir los resultados de búsqueda precalculados en las partes en primer lugar y asignar resultados de búsqueda precalculados recientemente creados/calculados a la parte respectiva (si la subdivisión en partes se refleja por la estructura de base de datos que no tiene que ser el caso, como se ha indicado más arriba, la subdivisión puede ser una división puramente lógica según reglas dadas, como, por ejemplo, en el ejemplo de que los resultados de búsqueda precalculados son recomendaciones de viaje con precios - recomendaciones de viaje con precios de origen y destino particulares y agrupadas en intervalos de tiempo de fecha de partida como, por ejemplo, todas las recomendaciones de viaje con precios para un par de origen-destino particular que parte entre hoy y hoy + 30 días (parte D1), para el mismo par de origen-destino que parte entre hoy + 31 días y hoy + 60 días (parte D2), etc.). Los factores de correlación resultantes asociados a las partes de resultados de búsqueda precalculados se almacenan en la representación 30 de datos interna a través del gestor 31 de entrada.

- Componente 30 de representación interna de datos: este componente provee herramientas para construir, almacenar, actualizar y acceder a matrices de datos de control que representan los resultados de búsqueda calculados previamente almacenados en la plataforma 4 de búsqueda. La función principal del componente 30 de representación interna de datos es proveer un "espejo de datos de control" de los resultados de búsqueda calculados previamente almacenados en la plataforma 4 de búsqueda que sirve como base para analizar los resultados de búsqueda calculados previamente para decidir cuáles de los resultados de búsqueda calculados previamente van a recalcularse durante el siguiente ciclo de recálculo. Más precisamente, el componente 30 de representación interna de datos no contiene una copia uno a uno de los resultados de búsqueda calculados previamente tal como se almacenan en la plataforma 4 de búsqueda, sino una representación de datos de control apropiada que no tiene que incluir los resultados de búsqueda calculados previamente en sí mismos tal como se almacenan en la plataforma 4 de búsqueda, pero, por otro lado, incluye datos de control asociados a los resultados de búsqueda calculados previamente como, por ejemplo, los tiempos de su último recálculo y, en particular, los valores de tendencia de validez λ_i y los valores de tasa de validez instantánea $\hat{\lambda}_i$.

- Gestor 31 de entrada: este componente introduce datos de control de diversas fuentes que incluyen el evaluador 27 de tendencia de validez, el evaluador 28 de tasa de validez instantánea y el evaluador 29 de correlación. Además, el gestor 31 de entrada recibe datos de control adicionales usados para mantener el modelo probabilístico como, por ejemplo, informe de popularidad de una base de datos de popularidad o fuente de datos, medición de costes de recálculo de una base de datos de costes de recálculo o fuente de datos, mediciones de precisión iniciales de una base de datos de precisión inicial o fuente de datos, y/o señales de eventos en tiempo real de fuentes que indican eventos en tiempo real que influyen potencialmente en la validez de los resultados de búsqueda precalculados. Estos datos de control adicionales se introducen a través de la interfaz 32 que representa esquemáticamente la(s) conexión(es) del gestor 31 de entrada a cualquiera de las bases de datos o fuentes de datos mencionadas anteriormente. El gestor 31 de entrada convierte los datos de control entrantes en los formatos de datos apropiados y actualiza las matrices de datos de control correspondientes que representan los resultados de búsqueda calculados previamente tal como se almacenan por el componente 30 de representación interna de datos .

- Analizador 33: este componente calcula matrices de datos intermedios implícitas en el modelo probabilístico (es decir, las probabilidades de validez de los resultados de búsqueda precalculados derivados de los datos de control del modelo probabilístico como, por ejemplo, edad, tendencias de validez, popularidad, precisión inicial) en base a las matrices almacenadas por el componente 30 de representación interna de datos.

- Gestor 34 de eventos: este componente agrega información sobre información de eventos en tiempo real y modifica las predicciones de validez dadas por el modelo probabilístico en consecuencia. Además, este componente se mejora para reconocer eventos asíncronos no señalizados externamente a través de la interfaz 32 en base a los valores de tasa de validez instantánea $\hat{\lambda}_i$. Para este fin, el gestor 34 de eventos lleva a cabo la comparación entre los valores de tasa de validez instantánea $\hat{\lambda}_i$ y los valores de tendencia de validez λ_i de partes correlacionadas con los resultados de búsqueda recalculados recientemente y corrige los parámetros del modelo probabilístico (las probabilidades de validez) si la comparación indica que los valores de tasa de validez instantánea $\hat{\lambda}_i$ están demasiado alejados de los valores de tendencia de validez λ_i .

- Optimizador 35: este componente ejecuta la estrategia de recálculo, por ejemplo, un recálculo orientado a la frecuencia de recálculo y la selección iterativa de resultados de búsqueda precalculados teniendo en cuenta los costes de cálculo variables de resultados de búsqueda precalculados interrelacionados como se describe en detalle en la solicitud europea 14290040.6. Después de haber determinado los resultados de búsqueda calculados previamente que van a recalcularse, el optimizador 35 genera órdenes de recálculo y emite estas órdenes de recálculo a la plataforma 3 de cálculo a través de la interfaz 36. Además, el optimizador 35 actualiza el tiempo de recálculo de estos resultados de búsqueda precalculados almacenados en el componente de representación 30 interna de datos .

- El módulo 37 de evaluación de validez es un complemento que genera estadísticas de la validez de los resultados de búsqueda precalculados a lo largo del tiempo basándose en los parámetros del modelo probabilístico que incluye los valores de tendencia de validez λ_i . Las estadísticas se envían a una pantalla de evaluación externa y/o para su presentación a un usuario.

Finalmente, la Fig. 13 es una representación esquemática de la estructura interna de un ordenador o servidor 120 que implementa los mecanismos para recalcular resultados de búsqueda precalculados como se describe en la presente memoria. El ordenador o servidor 120 está dispuesto para ejecutar un conjunto de instrucciones, para hacer que lleve a cabo cualquiera de las metodologías explicadas más arriba. El ordenador o servidor 120 incluye un procesador 121, una memoria 122 principal y, opcionalmente, una interfaz 123 de red inalámbrica (como, por ejemplo, una interfaz Wi-Fi y/o Bluetooth) y/o un dispositivo de interfaz de red móvil 2G/3G/4G, todos los cuales se comunican entre sí a través de un bus 124. Incluye además una memoria 125 estática, p. ej., una unidad de estado sólido y/o flash no extraíble y/o una tarjeta Micro o Mini SD extraíble, que almacena permanentemente el software que permite que el ordenador/servidor 120 ejecute las funciones del controlador 2 de recálculo como, por ejemplo, determinar los valores de tendencia de validez, los valores de tasa de validez instantánea, comparar ambos, adaptar las probabilidades de validez de los resultados de búsqueda precalculados correlacionados, seleccionar los resultados de búsqueda precalculados para el recálculo, etc. y opcionalmente comunicarse con ordenadores/dispositivos cliente dentro de una red de área local o amplia a través de su dispositivo 123 de interfaz de red por cable y/o inalámbrico. Además, el ordenador/servidor 120 incluye una pantalla 127, un módulo 129 de control de interfaz de usuario y un dispositivo 128 alfanumérico y de entrada de cursor. Opcionalmente, pueden estar presentes interfaces 126 de E/S adicionales como, por ejemplo, lector de tarjetas e interfaces USB. Un conjunto ejecutable de instrucciones (es decir, software) 130 que incorpora cualquiera, o todas, de las metodologías descritas más arriba, reside completamente, o al menos parcialmente, permanentemente en la memoria 125 permanente. Cuando se ejecutan, los datos de proceso respectivos residen en la memoria 122 principal y/o procesador 121. El software 130 puede transmitirse o recibirse además como una señal 132 propagada a través del dispositivo 123 de interfaz de red por cable o inalámbrico de/a un servidor de software dentro de la red de área local o Internet.

Aunque en la presente memoria se han descrito ciertos productos y métodos contruidos según las enseñanzas de la invención, el alcance de cobertura de esta patente no se limita a los mismos. Por el contrario, esta patente cubre todas las realizaciones de las enseñanzas de la invención que caen bastante dentro del alcance de las reivindicaciones anexas, ya sea literalmente o bajo la doctrina de equivalentes.

REIVINDICACIONES

1. Un método de gestión de resultados de búsqueda precalculados, llevándose a cabo el método en un entorno (1) de base de datos, comprendiendo el entorno de base de datos:

5 - al menos una plataforma (4) de búsqueda para mantener resultados de búsqueda precalculados, estando los resultados de búsqueda precalculados subdivididos en múltiples partes de resultados de búsqueda precalculados relacionados que incluyen una primera parte D;

- un controlador (2) de recálculo para controlar el recálculo de los resultados de búsqueda precalculados basándose en probabilidades de validez que se asocian a los resultados de búsqueda precalculados; y

- una plataforma (3) de cálculo para recalcular los resultados de búsqueda precalculados;

10 comprendiendo el método:

- determinar (12), por el controlador (2) de recálculo, una tendencia de validez λ_i que indica una tasa de cambio de la probabilidad de validez de un resultado de búsqueda i precalculado a lo largo del tiempo, siendo el resultado de búsqueda i precalculado un miembro de la primera parte D, siendo la tendencia de validez λ_i derivada de al menos tres recálculos pasados de i ;

15 - determinar (13), por el controlador de recálculo, una tasa de validez instantánea $\hat{\lambda}_i$ para el resultado de búsqueda i precalculado, la tasa de validez instantánea $\hat{\lambda}_i$ siendo derivada a partir de los dos últimos recálculos de los resultados de búsqueda precalculados relacionados incluidos en la primera parte D, en donde la tasa de validez instantánea $\hat{\lambda}_i$ se determina como un valor común $\hat{\lambda}_D$ para los resultados de búsqueda precalculados en la primera parte D;

20 - en respuesta a la determinación (14) de que una diferencia entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i supera un grado dado con la tasa de validez instantánea $\hat{\lambda}_i$ siendo mayor que la tendencia de validez λ_i disminuir (15), por el controlador (2) de recálculo, las probabilidades de validez asociadas a otros resultados de búsqueda precalculados incluidos en una segunda parte D' que se correlacionan con la primera parte D, en donde la cantidad de disminución depende de la cantidad de la diferencia entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i ;

25 - en donde los resultados de búsqueda precalculados relacionados dentro de una parte tienen tendencias de validez λ_i idénticas o similares;

- en donde la correlación entre la primera parte D y la segunda parte D' viene dada por $\frac{\sigma_{DD'}}{\sigma_D \sigma_{D'}}$ que supera un umbral dado, en donde $\sigma_{DD'}$ denota una covarianza entre valores pasados de la tasa de validez instantánea $\hat{\lambda}_{iD}$ de la primera parte D y de la tasa de validez instantánea $\hat{\lambda}_{iD'}$ de la segunda parte D' en el tiempo, σ_D^2 denota una varianza de los valores pasados de la tasa de validez instantánea $\hat{\lambda}_{iD}$ en el tiempo y por lo tanto σ_D denota una desviación estándar de la tasa de validez instantánea $\hat{\lambda}_{iD}$ de la primera parte D y $\sigma_{D'}$ denota una desviación estándar de la tasa de validez instantánea $\hat{\lambda}_{iD'}$ de la segunda parte D', y

35 - emitir (16), por el controlador (2) de recálculo, órdenes de recálculo a la plataforma (3) de cálculo para recalcular una parte de los resultados de búsqueda precalculados con probabilidades de validez más bajas que otros resultados de búsqueda precalculados con probabilidades de validez más altas.

2. El método de la reivindicación 1, en donde las probabilidades de validez que se asocian a los otros resultados de búsqueda precalculados incluidos en la segunda parte D' se disminuyen en dependencia de un grado de correlación entre la primera parte D y la segunda parte D'.

40 3. El método de la reivindicación 2, en donde el grado de correlación entre la primera parte D y la segunda parte D' se determina mediante una regresión lineal para modelar una relación entre

- la diferencia entre una tendencia de validez λ_D y la tasa de validez instantánea $\hat{\lambda}_{iD}$, λ_D indicando una tasa de cambio de la probabilidad de validez de los resultados de búsqueda precalculados en la primera parte D; y

- una diferencia entre una tendencia de validez $\lambda_{D'}$ y la tasa de validez instantánea $\hat{\lambda}_{iD'}$, $\lambda_{D'}$ indicando una tasa de cambio de la probabilidad de validez de los resultados de búsqueda precalculados en la segunda parte D'.

45

4. El método de la reivindicación 3, en donde el grado de correlación entre la primera parte D y la segunda parte D' viene dado por la relación

$$\widehat{\lambda}_{D'} - \lambda_{D'} \approx \frac{\sigma_{DD'}}{\sigma_D^2} (\widehat{\lambda}_D - \lambda_D).$$

5. El método de cualquiera de las reivindicaciones 1 a 4, en donde la probabilidad de validez asociada al resultado de búsqueda i precalculado viene dada por una probabilidad de que el resultado de búsqueda precalculado siga siendo válido después de un tiempo T_n definido por $e^{-\lambda_i T_n}$, en donde al menos uno de determinar la tendencia de validez λ_i y determinar la tasa de validez instantánea $\hat{\lambda}_i$ comprende una estimación de probabilidad máxima para estimar un valor esperado para λ_i dentro de $e^{-\lambda_i T_n}$ y $\hat{\lambda}_i$ dentro de $e^{-\hat{\lambda}_i T_n}$, respectivamente, la estimación de probabilidad máxima basándose en una serie de muestras de pares de recálculo, incluyendo cada par de recálculo un tiempo desde el último recálculo de i y un indicador que indica si i ha cambiado o no desde el último recálculo de i.

6. El método de la reivindicación 5, que comprende además calcular un valor de confianza para la tendencia de validez estimada λ_i y/o la tasa de validez instantánea estimada $\hat{\lambda}_i$ usando la estimación de probabilidad máxima.

7. El método de cualquiera de las reivindicaciones 1 a 4, en donde la determinación de la tendencia de validez λ_i y/o la determinación de la tasa de validez instantánea $\hat{\lambda}_i$ comprende establecer una función de distribución empírica de la probabilidad de que el resultado de búsqueda i precalculado permanezca válido a lo largo del tiempo y hacer una regresión de la función de distribución empírica para ajustarse a una función modelo.

8. El método de la reivindicación 7, en donde la función de distribución empírica para la tendencia de validez λ_i y/o la tasa de validez instantánea $\hat{\lambda}_i$ se revierte para ajustar una función exponencial de la forma $f(t) = e^{-\lambda_i t}$ y/o $f(t) = e^{-\hat{\lambda}_i t}$.

9. El método de cualquiera de las reivindicaciones 1 a 8, en donde al menos uno de determinar la tendencia de validez λ_i y determinar la tasa de validez instantánea $\hat{\lambda}_i$ utiliza procesos MapReduce.

10. El método de cualquiera de las reivindicaciones 1 a 9, en donde la tasa de validez instantánea $\hat{\lambda}_i$ para el resultado de búsqueda i precalculado se determina en respuesta a un recálculo de los resultados de búsqueda precalculados incluidos en la primera parte D.

11. Un controlador (2) de recálculo para gestionar resultados de búsqueda precalculados mantenidos por una plataforma (4) de búsqueda, subdividiéndose los resultados de búsqueda precalculados en múltiples partes de resultados de búsqueda precalculados relacionados que incluyen una primera parte D, comprendiendo el controlador (2) de recálculo uno o más procesadores (101) y una memoria (102, 106, 111) que incluye instrucciones que, cuando se ejecutan por el uno o más procesadores (101), hacen que el controlador (2) de recálculo controle el recálculo de los resultados de búsqueda precalculados basándose en probabilidades de validez que se asocian a los resultados de búsqueda precalculados mediante:

- determinación (12) de una tendencia de validez λ_i que indica una tasa de cambio de la probabilidad de validez del resultado de búsqueda i precalculado a lo largo del tiempo, siendo el resultado de búsqueda i precalculado un miembro de la primera parte D, siendo la tendencia de validez λ_i derivada de al menos tres recálculos pasados de i;

- determinación (13) de una tasa de validez instantánea $\hat{\lambda}_i$ para el resultado de búsqueda i precalculado, siendo la tasa de validez instantánea $\hat{\lambda}_i$ derivada de los dos últimos recálculos de los resultados de búsqueda precalculados relacionados incluidos en la primera parte D, en donde la tasa de validez instantánea $\hat{\lambda}_i$ se determina como un valor común $\hat{\lambda}_D$ para los resultados de búsqueda precalculados en la primera parte D;

- en respuesta a la determinación (14) de una diferencia entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i supera un grado dado con la tasa de validez instantánea $\hat{\lambda}_i$ siendo mayor que la tendencia de validez λ_i , disminuir (15) probabilidades de validez asociadas a otros resultados de búsqueda precalculados incluidos en una segunda parte D' correlacionada con la primera parte D, en donde la cantidad de disminución depende de la cantidad de la diferencia entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i ;

- en donde los resultados de búsqueda precalculados relacionados dentro de una parte tienen tendencias de validez λ_i idénticas o similares;

- en donde la correlación entre la primera parte D y la segunda parte D' viene dada por $\frac{\sigma_{DD'}}{\sigma_D \sigma_{D'}}$ que supera un umbral dado, en donde $\sigma_{DD'}$ denota una covarianza entre valores pasados de la tasa de validez instantánea $\hat{\lambda}_i$ de la primera

parte D y de la tasa de validez instantánea $\hat{\lambda}_{D'}$ de la segunda parte D' en el tiempo, σ_D^2 denota una varianza de los valores pasados de la tasa de validez instantánea $\hat{\lambda}_D$ en el tiempo y por lo tanto σ_D denota una desviación estándar de la tasa de validez instantánea $\hat{\lambda}_D$ de la primera parte D y $\sigma_{D'}$ denota una desviación estándar de la tasa de validez instantánea $\hat{\lambda}_{D'}$ de la segunda parte D', y

5 - emisión (16) de órdenes de recálculo a una plataforma de cálculo para recalcular una parte de los resultados de búsqueda precalculados con probabilidades de validez más bajas que otros resultados de búsqueda precalculados con probabilidades de validez más altas.

10 12. El controlador de recálculo de la reivindicación 11, en donde las instrucciones, cuando se ejecutan por el uno o más procesadores, hacen que el controlador de recálculo lleve a cabo los métodos de cualquiera de las reivindicaciones 2 a 10.

13. Un producto de programa informático para gestionar resultados de búsqueda precalculados mantenidos por una plataforma (4) de búsqueda, subdividiéndose los resultados de búsqueda precalculados en múltiples partes de resultados de búsqueda precalculados relacionados que incluyen una primera parte D, comprendiendo el producto de programa informático:

15 un medio de almacenamiento legible por ordenador no transitorio; e

instrucciones almacenadas en el medio de almacenamiento legible por ordenador no transitorio que, cuando se ejecutan por un procesador (101), hacen que el procesador (101) controle el recálculo de los resultados de búsqueda precalculados basándose en probabilidades de validez que se asocian a los resultados de búsqueda precalculados mediante:

20 - determinación (12) de una tendencia de validez λ_i que indica una tasa de cambio de la probabilidad de validez de un resultado de búsqueda i precalculado a lo largo del tiempo, siendo el resultado de búsqueda i precalculado un miembro de la primera parte D, la tendencia de validez λ_i siendo derivada de al menos tres recálculos pasados de i;

25 - determinación (13) de una tasa de validez instantánea $\hat{\lambda}_i$ para el resultado de búsqueda i precalculado, la tasa de validez instantánea $\hat{\lambda}_i$ derivándose de los dos últimos recálculos de resultados de búsqueda precalculados relacionados incluidos en la primera parte D, en donde la tasa de validez instantánea $\hat{\lambda}_i$ se determina como un valor común $\hat{\lambda}_D$ para los resultados de búsqueda precalculados en la primera parte D;

30 - en respuesta a la determinación (14) de que una diferencia entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i supera un grado dado con la tasa de validez instantánea $\hat{\lambda}_i$ siendo mayor que la tendencia de validez λ_i , reducir (15) probabilidades de validez asociadas a otros resultados de búsqueda precalculados incluidos en una segunda parte D' correlacionada con la primera parte D, en donde la cantidad de disminución depende de la cantidad de la diferencia entre la tasa de validez instantánea $\hat{\lambda}_i$ y la tendencia de validez λ_i ;

- en donde los resultados de búsqueda precalculados relacionados dentro de una parte tienen tendencias de validez λ_i idénticas o similares;

35 - en donde la correlación entre la primera parte D y la segunda parte D' viene dada por $\frac{\sigma_{DD'}}{\sigma_D \sigma_{D'}}$ que supera un umbral dado, en donde $\sigma_{DD'}$ denota una covarianza entre valores pasados de la tasa de validez instantánea $\hat{\lambda}_D$ de la primera parte D y de la tasa de validez instantánea $\hat{\lambda}_{D'}$ de la segunda parte D' en el tiempo, σ_D^2 denota una varianza de los valores pasados de la tasa de validez instantánea $\hat{\lambda}_D$ en el tiempo y por lo tanto σ_D denota una desviación estándar de la tasa de validez instantánea $\hat{\lambda}_D$ de la primera parte D y $\sigma_{D'}$ denota una desviación estándar de la tasa de validez instantánea $\hat{\lambda}_{D'}$ de la segunda parte D', y

40 - emisión (16) de órdenes de recálculo a una plataforma de cálculo para recalcular una parte de los resultados de búsqueda precalculados con probabilidades de validez más bajas que los resultados de búsqueda precalculados con probabilidades de validez más altas.

45 14. El producto de programa informático de la reivindicación 13, en donde las instrucciones, cuando son ejecutadas por el uno o más procesadores, hacen que el procesador lleve a cabo los métodos de cualquiera de las reivindicaciones 2 a 10.

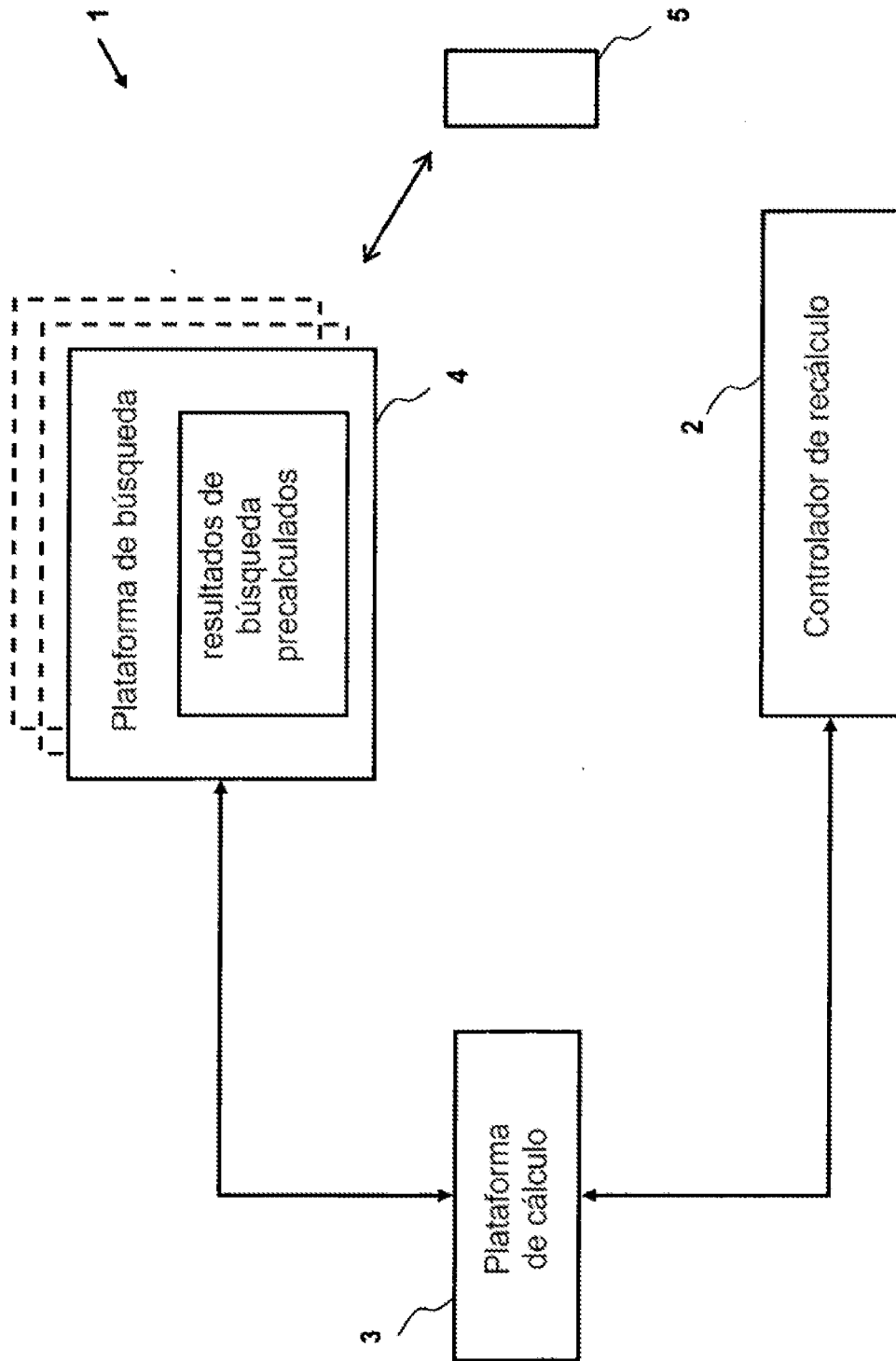


Fig. 1

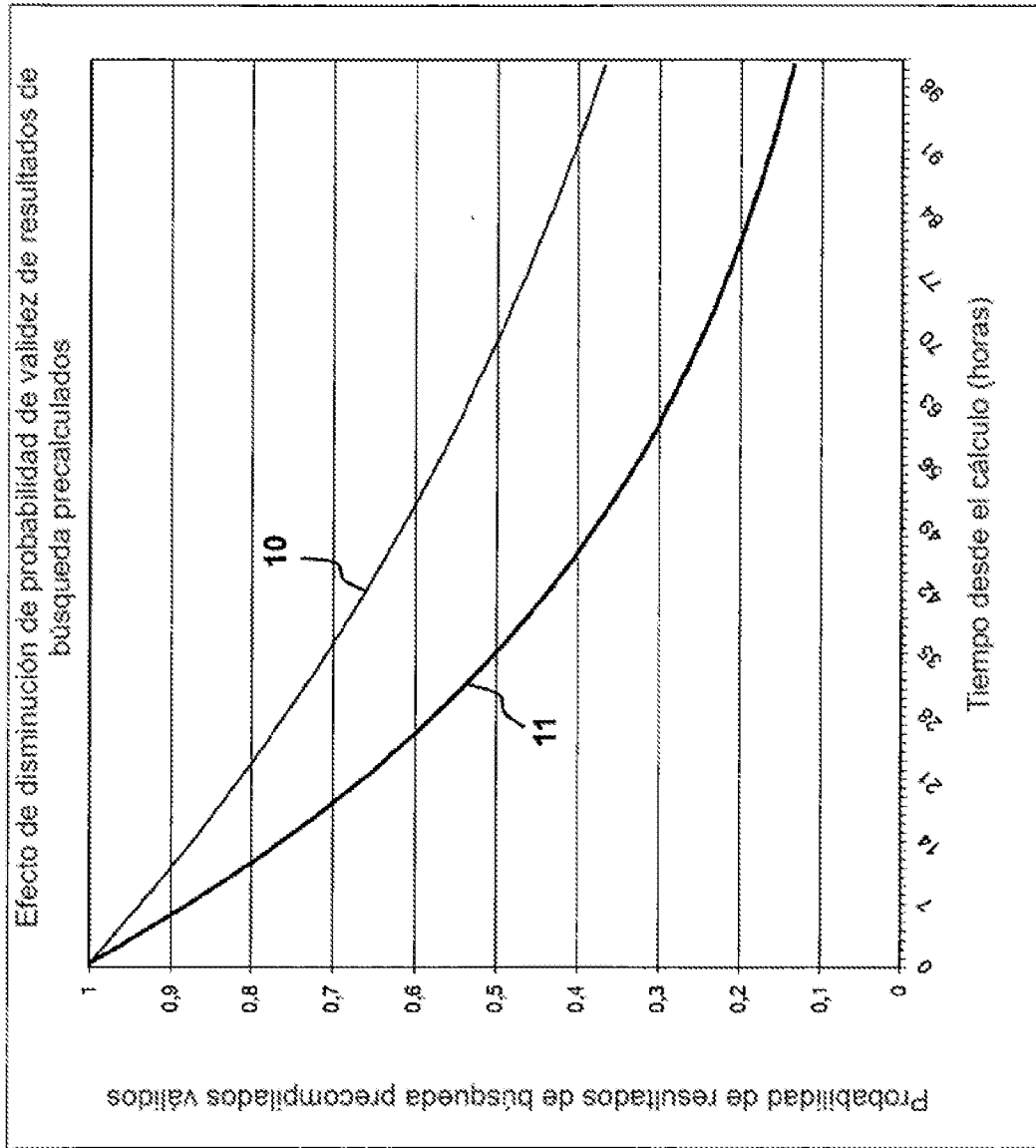


Fig. 2

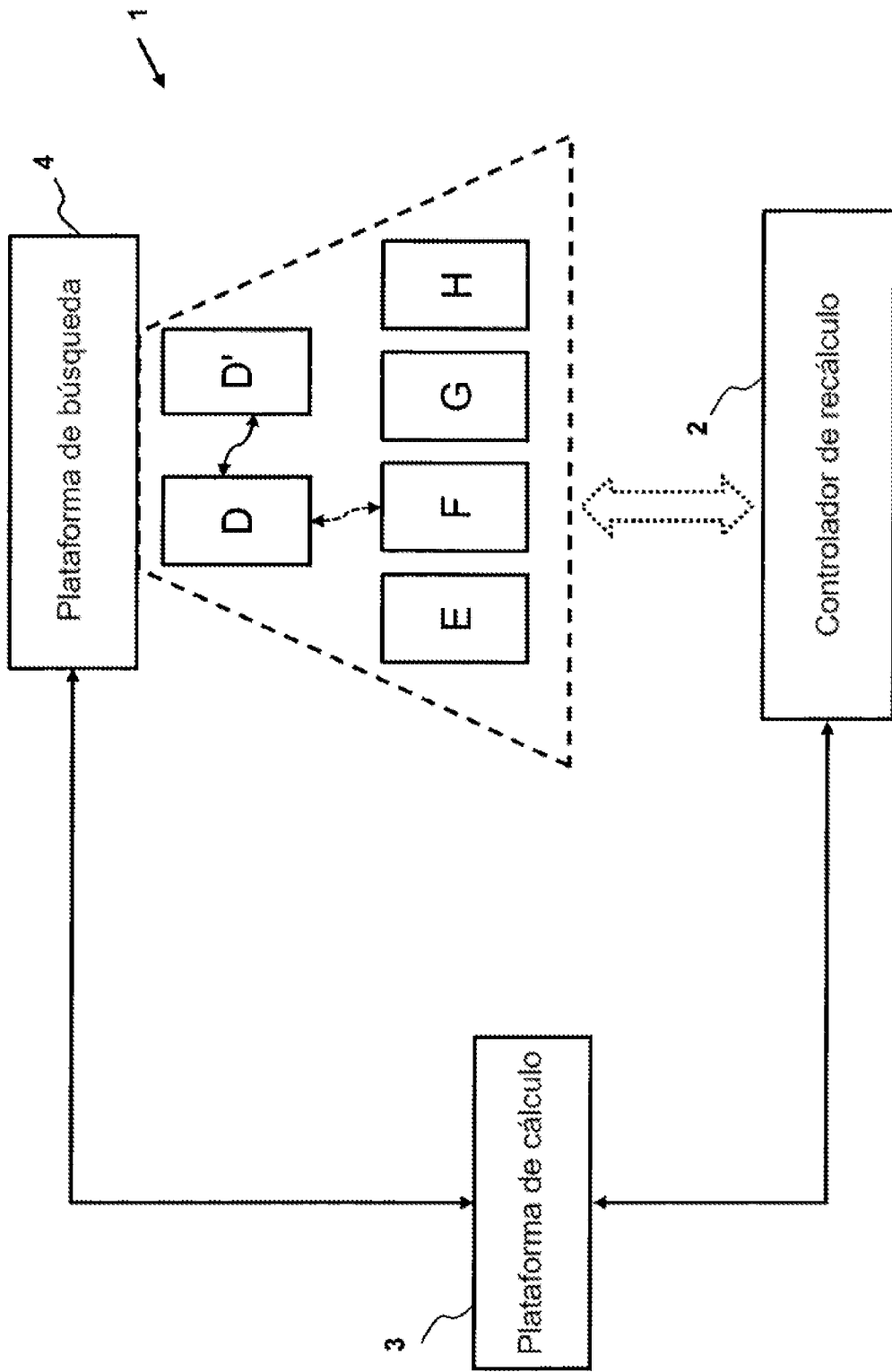


Fig. 3

Recálculos sucesivos de un resultado de búsqueda i precalculado:

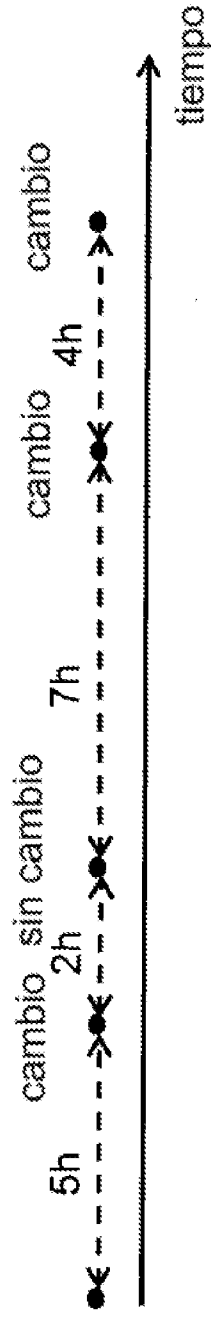


Fig. 4a

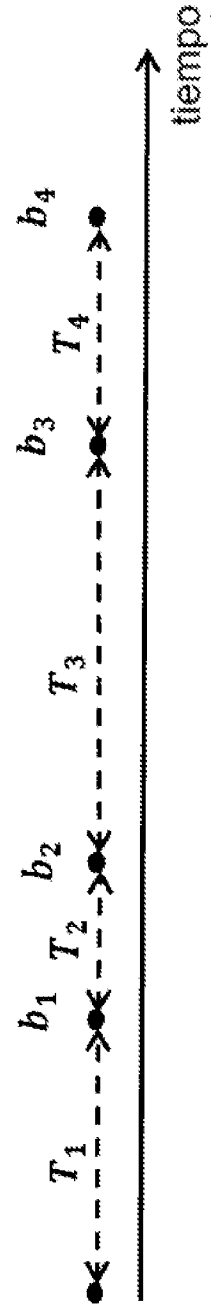


Fig. 4b

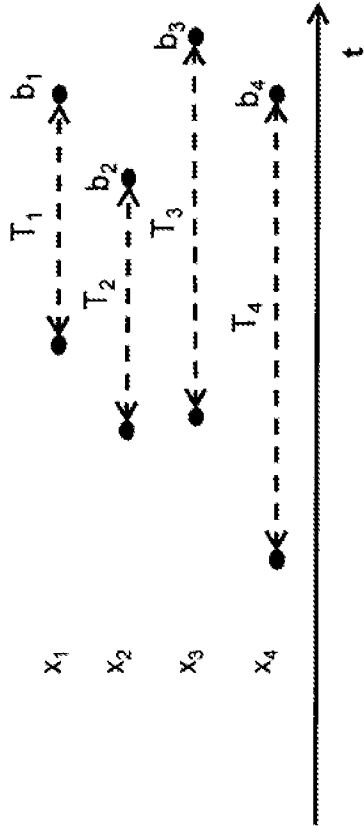


Fig. 5a

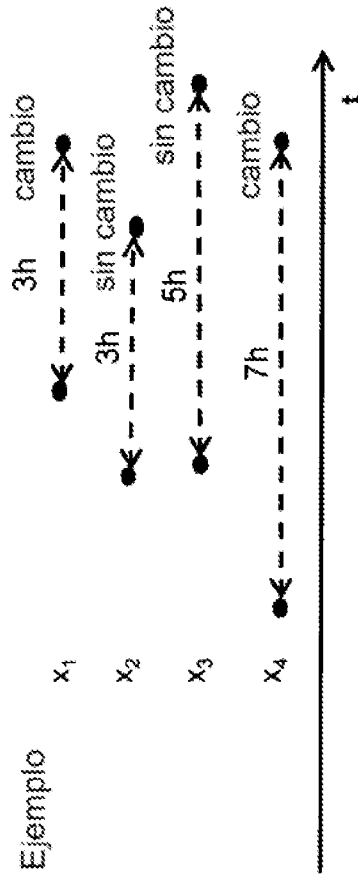


Fig. 5b

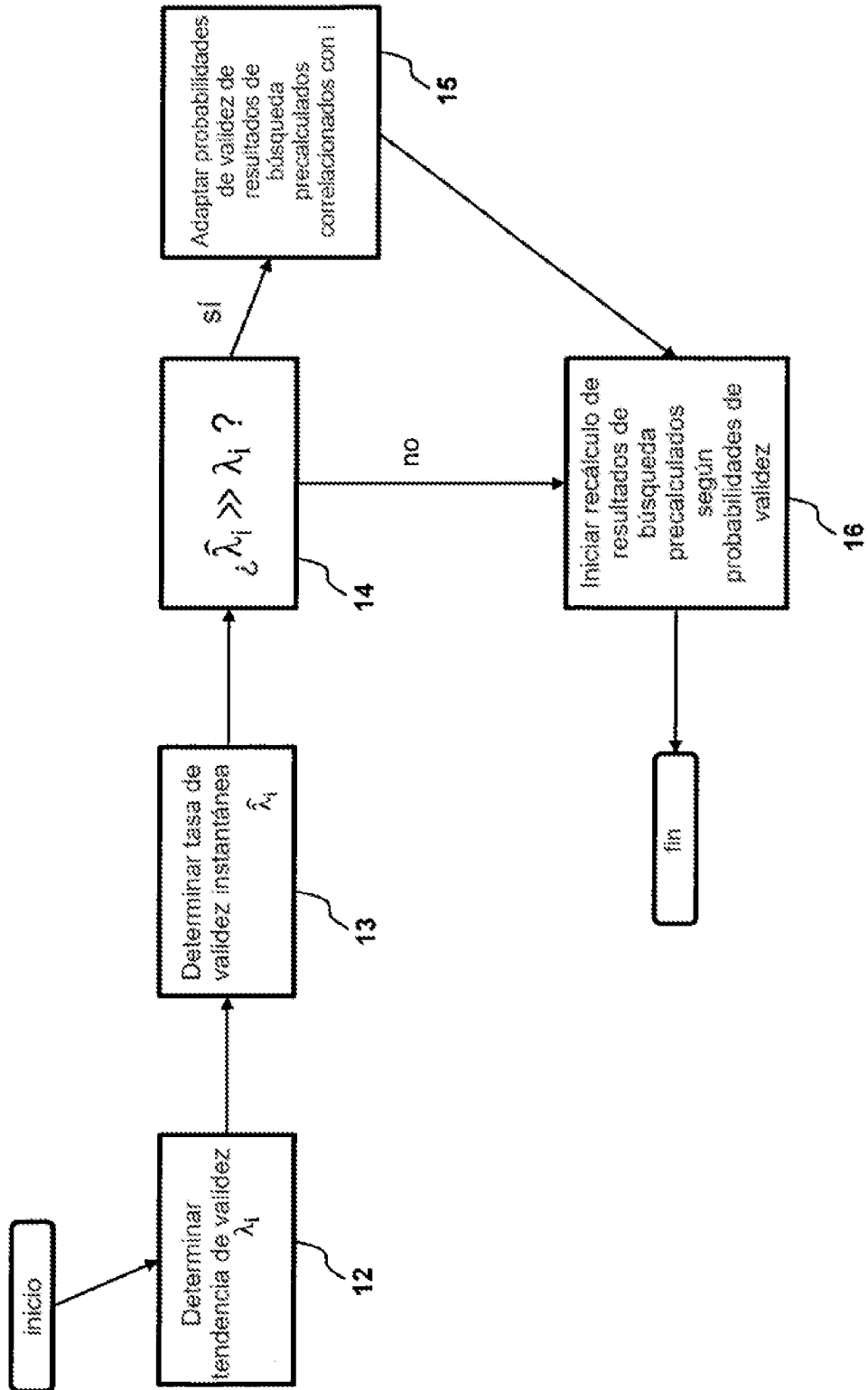


Fig. 6

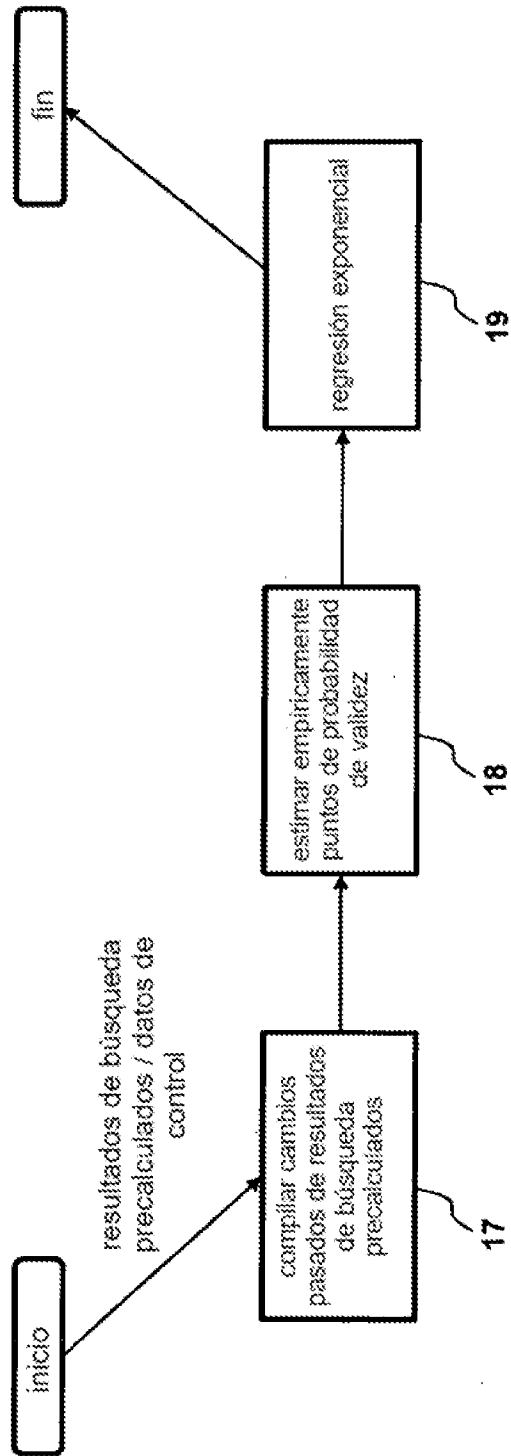


Fig. 7

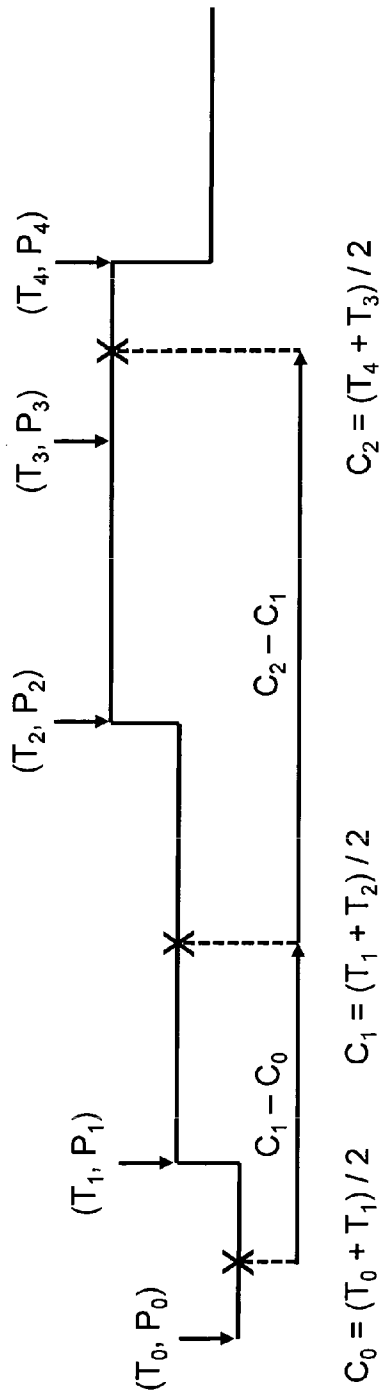


Fig. 8a

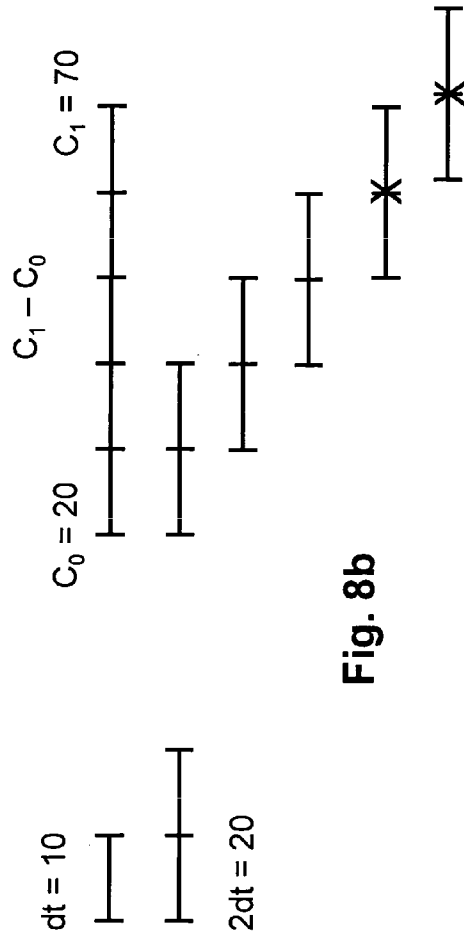
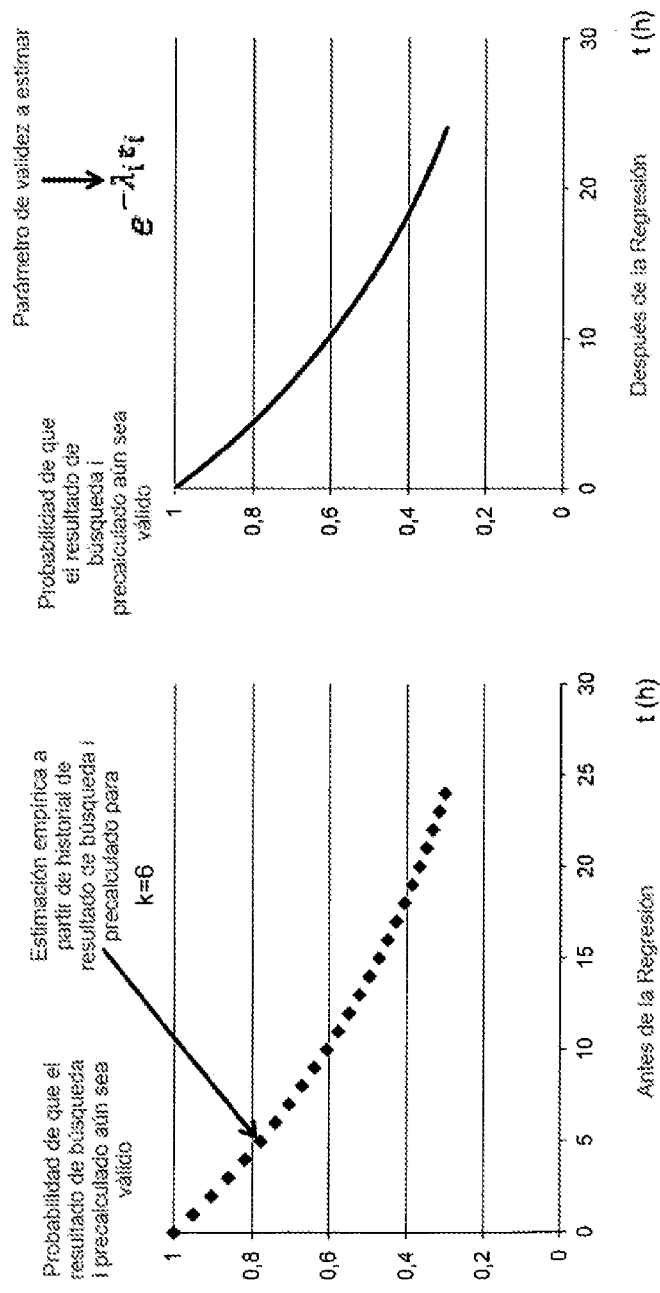


Fig. 8b



$\lambda_i = 0,05/h$ Fig. 9b

Fig. 9a

Probabilidades de validez según tendencia de validez (modelo probabilístico) y tasa de validez instantánea

| Parte | A | B | C | D | E | F |
|--------------|------------------------------|------------------------------|------------------------------|--|------------------------------|-------------|
| Fechas part. | JUL | AGO | SEP | OCT | NOV | DIC |
| 1 PAR-NYC | Modelo: 85% Instante: 85% | Modelo: 87% Instante: 85% | Modelo: 90% | Modelo: 93% | Modelo: 95% Instante: 90% | Modelo: 99% |
| 2 NCE-LON | Modelo: 82% | Modelo: 84% | Modelo: 85% | Modelo: 85% <u>Instante: 62%</u> | Modelo: 90% | Modelo: 98% |
| 3 BOS-MIA | Modelo: 80% | Modelo: 91% Instante: 95% | Modelo: 89% Instante: 84% | Modelo: 95% | Modelo: 97% | Modelo: 97% |

Fig. 10

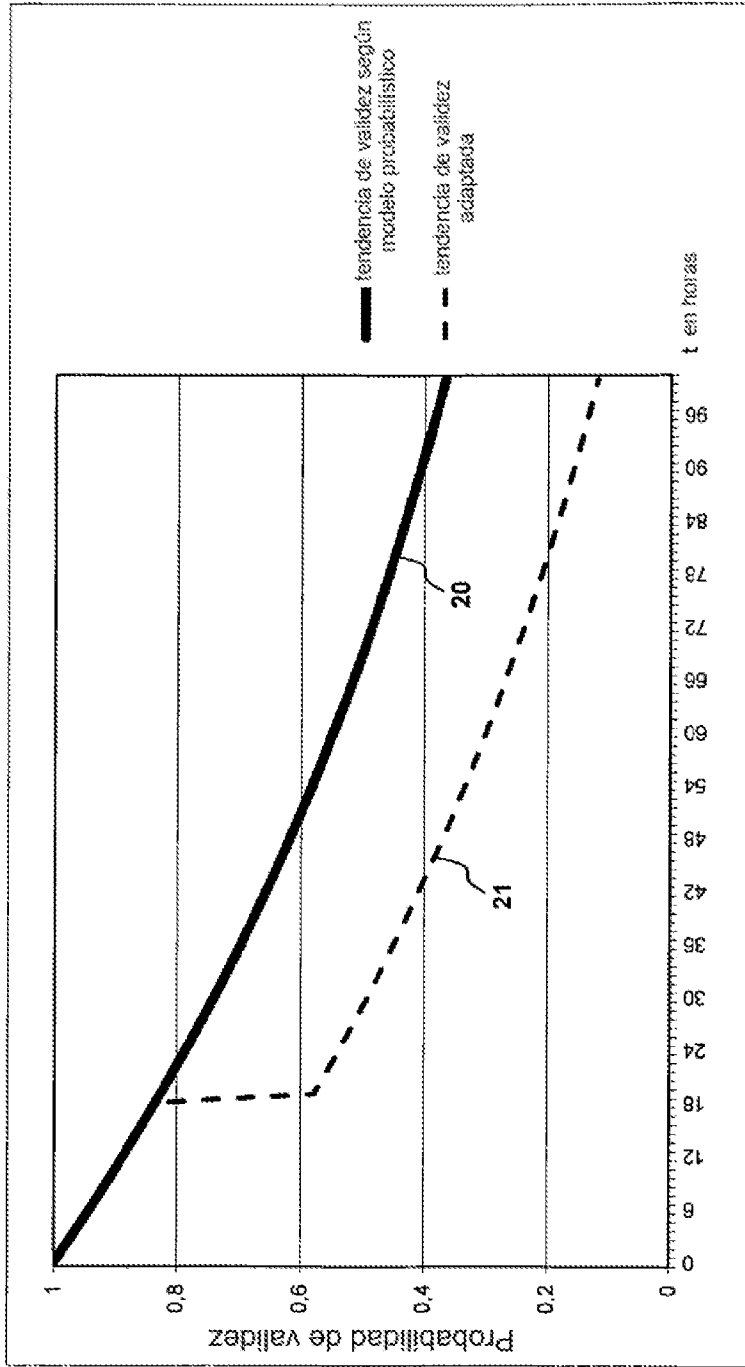


Fig. 11

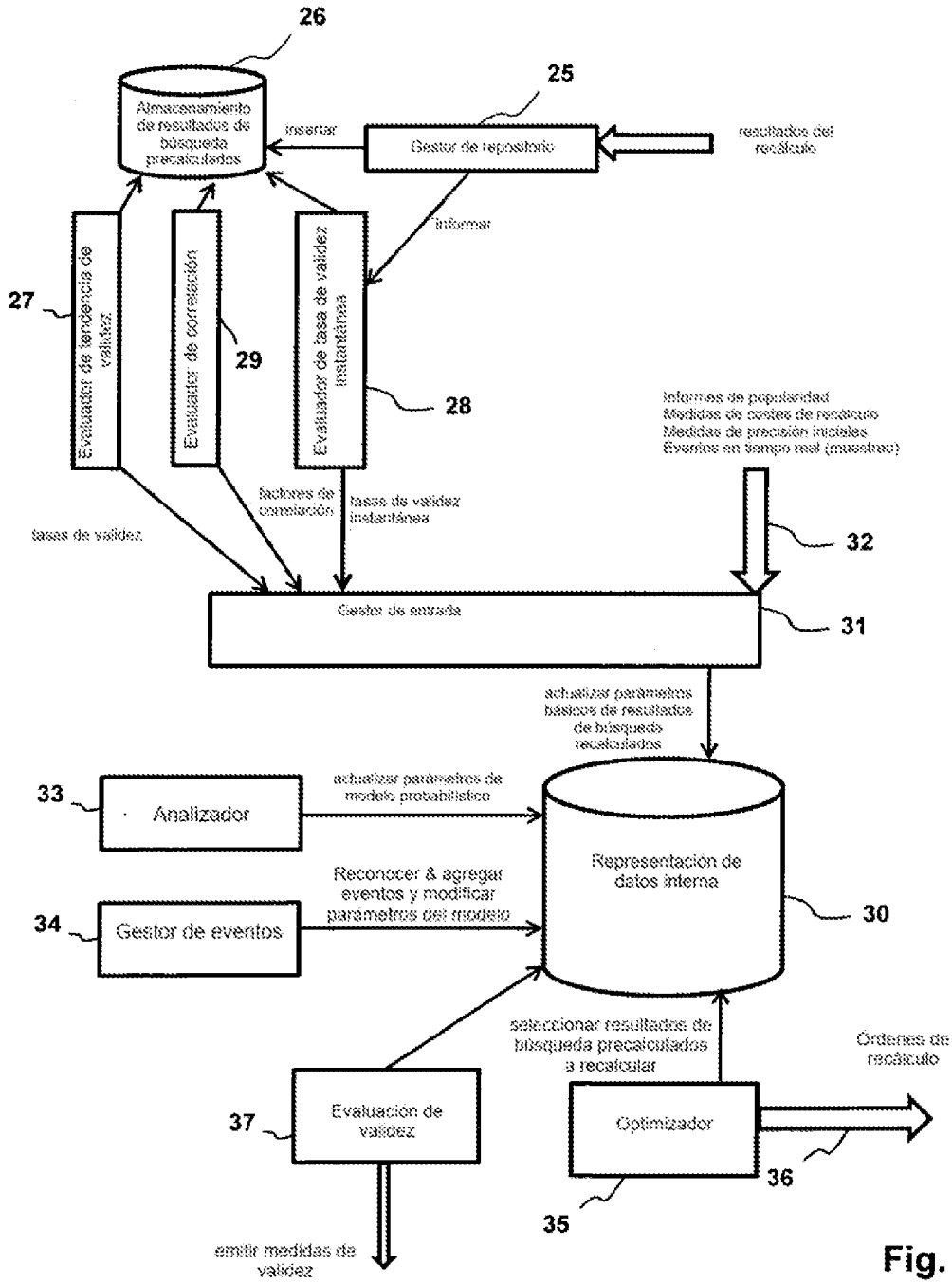


Fig. 12

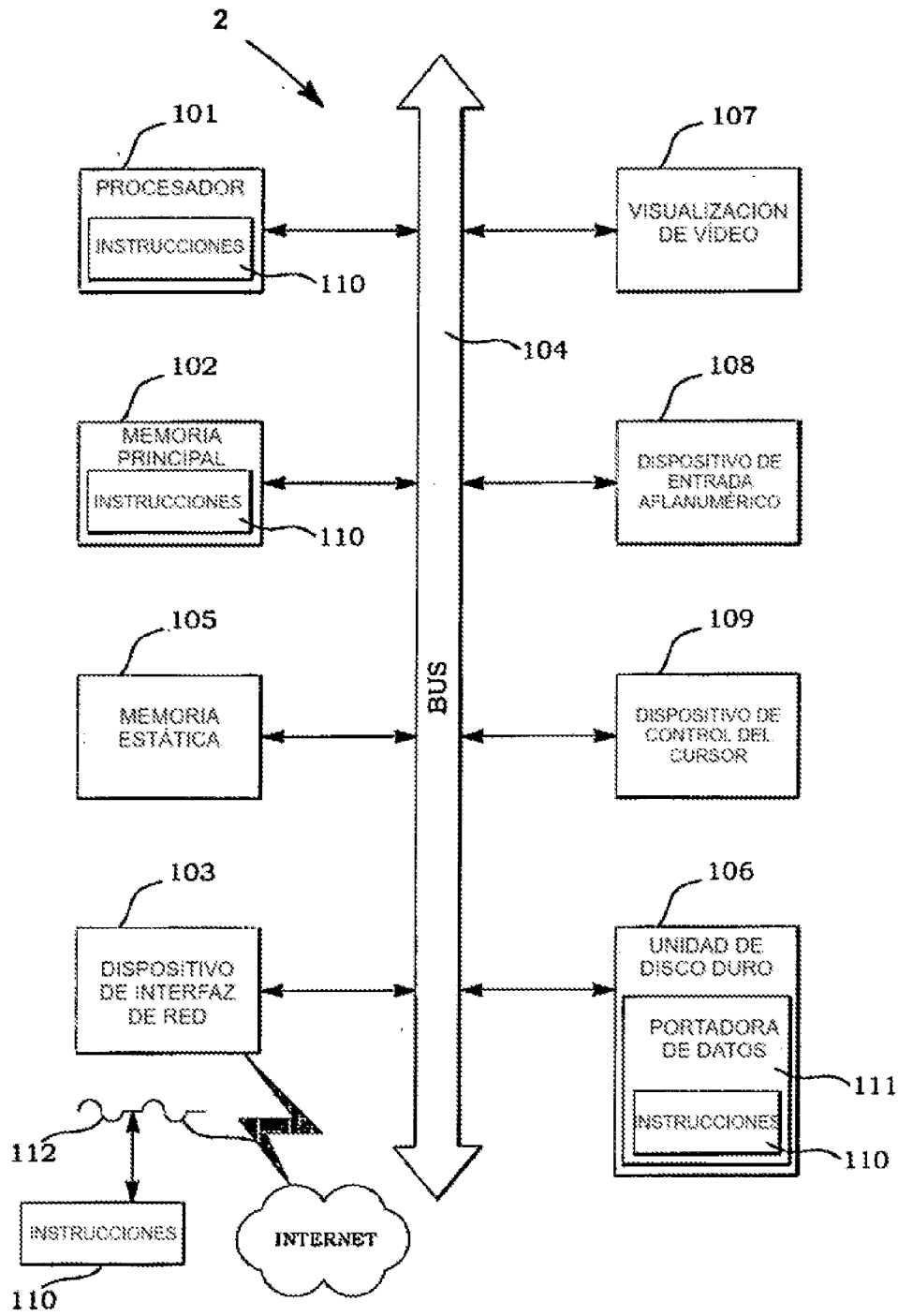


Fig. 13