

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 622 117**

51 Int. Cl.:

G06F 11/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **04.12.2007** **E 07023415 (8)**

97 Fecha y número de publicación de la concesión europea: **22.02.2017** **EP 2071461**

54 Título: **Método y sistema para estimar un número de usuarios de un sitio web en base a datos comprimidos con pérdidas**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.07.2017

73 Titular/es:

**COREMEDIA AG (100.0%)
LUDWIG-ERHARD-STRASSE 18
20459 HAMBURG, DE**

72 Inventor/es:

KUMMER, OLAF

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 622 117 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

Método y sistema para estimar un número de usuarios de un sitio web en base a datos comprimidos con pérdidas

5 La presente invención se refiere a un método y sistema para estimar un número de usuarios de un sitio web.

El número de usuarios de un sitio web es un parámetro importante para determinar el éxito de la página web y a menudo tiene una influencia directa en los ingresos por publicidad.

10 En la técnica anterior se conocen métodos para identificar entidades que acceden a una página web. En general se usa la dirección IP de la entidad accedente o la entidad es identificada usando un cookie. Además, algunas páginas web requieren que el usuario se registre introduciendo su nombre de entrada y una contraseña. En este caso, la entidad accedente puede ser identificada fácilmente mediante el nombre de usuario.

15 A este respecto es importante entender que es casi imposible estar seguro acerca del usuario que envía una petición que tiene una dirección IP, que usa un ordenador en el que se almacena un cookie, o que se registra en un sitio web. La dirección IP puede ser reescrita por un cortafuegos, el ordenador puede ser usado por más de una persona, o alguien puede usar el nombre de usuario de otra persona. No obstante, las entidades identificadas se utilizan a menudo para estimar el número de usuarios de un sitio web porque se considera que son un proxy
20 razonable para los usuarios.

Una vez que la entidad es identificada, se almacenan registros que reflejan el acceso de la entidad. Al final de un intervalo, una herramienta de análisis lee los registros almacenados y cuenta el número de entidades accedentes que se considera que es una buena estimación del número de usuarios accedentes.

25 Un problema de este procedimiento es que, hasta el final del intervalo de reporte, todos los registros tienen que ser almacenados. Para pequeños sitios web, este inconveniente es aceptable. Sin embargo, para sitios web grandes, la cantidad de datos almacenados es excesiva.

30 Para hacer frente a este problema, en la técnica anterior se ha propuesto el muestreo. El principio básico del muestreo es que solamente por cada x-ésimo (por ejemplo, cada 10-ésimo) acceso se almacena la identidad de una entidad accedente. Al final del intervalo de reporte, se analizan los registros almacenados y se estima un número de entidades accedentes en base a los registros muestreados.

35 Sin embargo, este método de muestreo tiene el inconveniente de que sólo es preciso cuando el número de eventos, por ejemplo, accesos registrados, en el intervalo de reporte es grande. Sin embargo, mientras que los intervalos largos, como años completos, contienen típicamente eventos suficientes, esto podría no ser verdadero con respecto a intervalos horarios o diarios.

40 Además, si el porcentaje de muestreo se establece antes de muestrear los datos, hay el peligro de que el porcentaje de muestreo elegido sea demasiado bajo, dando lugar a inexactitudes, o demasiado alto, degradando por ello el rendimiento.

45 US 2006/274763 A1 muestra un método de muestreo según el que un conjunto de datos conteniendo datos de tráfico de sitio web u otros datos es muestreado según una tasa de muestreo variable. Se establece un número deseado de muestras por período de tiempo, y se determina una tasa de muestreo base. Los elementos de datos en el conjunto de datos son muestreados según la tasa de muestreo base, para obtener un conjunto de muestras. Para períodos de tiempo donde el tamaño del conjunto de muestras resultante excede del número deseado de muestras, se establece una nueva tasa de muestreo y los elementos de datos para el período de tiempo son muestreados de nuevo. Consiguientemente, hay que realizar un re-muestreo engorroso según la idea de US 2006/274763 A1.
50

US 2006/277197 A1 muestra un método según el que se facilita un formato de datos para almacenar datos tales como datos de tráfico de sitio web. El formato de datos descrito permite el acceso y la filtración de datos, por ejemplo al generar reportes de tráfico de sitio web. El formato de datos también proporciona compresión de datos.

55 Un método para generar un archivo de datos según el formato de datos emplea compresión lineal e indexación para almacenar los datos.

60 EP 1 341 089 A2 muestra un método para seguimiento y reporte de la actividad de tráfico en un sitio web, incluyendo el método almacenar una página web en un primer servidor acoplado a una red. Cuando la página web es pedida por un ordenador visitante, el ordenador visitante es seleccionado (o no es seleccionado) para inclusión dentro de un grupo de muestras, donde el grupo de muestras es solamente un subconjunto del tráfico total al sitio web. Se almacena un indicador de selección tal como un cookie en el ordenador visitante asociado con dicha selección. Un código de extracción de datos dentro del sitio web es operado o no es operado dependiendo del valor del indicador de selección. El código de extracción de datos es operado, y la actividad del ordenador visitante en el sitio web es rastreada en consecuencia solamente si el ordenador visitante se selecciona como un elemento dentro del grupo de muestras; de otro modo el código de extracción de datos no es operado y no se genera actividad de
65

tráfico.

5 En vista de los problemas antes descritos, un objeto de la invención es proporcionar un método y un sistema para estimar un número de usuarios de un sitio web que sean eficientes y que permitan estimar el número de usuarios con una precisión aceptable.

Este objeto se logra con la materia de las reivindicaciones independientes. Las reivindicaciones dependientes especifican realizaciones preferibles.

10 La invención incluye un método para estimar un número de usuarios de un sitio web que incluye los pasos de determinar repetidas veces que una entidad accede al sitio web, en reacción al acceso al sitio web, determinar datos dependientes de la entidad que accede al sitio web y almacenar dichos datos, comprimir con pérdidas dichos datos almacenados, y estimar un número de usuarios del sitio web en base a los datos almacenados comprimidos.

15 Esto quiere decir que, según la invención, por ejemplo, los datos acerca de la identidad de entidades accedentes son almacenados y dichos datos almacenados son comprimidos más tarde usando un algoritmo de compresión con pérdidas. En base a los datos comprimidos con pérdidas, es posible estimar un número de usuarios del sitio web.

20 Este método tiene la ventaja de que, debido al algoritmo de compresión con pérdidas, hay que almacenar y analizar menos datos. Una pequeña pérdida de precisión es aceptable, porque la determinación de datos dependiente de la entidad es ya una fuente de inexactitudes.

25 Específicamente, realizando dichos pasos de compresión con pérdidas y estimación, el método según la presente invención incluye los pasos de proporcionar una matriz de bits en un almacenamiento, mapear repetidas veces los datos dependientes de las entidades que acceden al sitio web sobre un bit de la matriz de bits según una función de mapeado y activar el bit de la matriz de bits sobre el que se mapearon los datos, y estimar el número de usuarios del sitio web en base a los bits activados de la matriz de bits.

30 Esto tiene la ventaja de que permite almacenar eficientemente datos que pueden ser usados para estimar un número de usuarios del sitio web. Preferiblemente, solamente se establecen bits únicos en respuesta a los accesos. De esta forma, la cantidad necesaria de almacenamiento es a menudo inferior en comparación con almacenar como números las identidades exactas de las entidades accedentes.

35 Preferiblemente, el método incluye los pasos de determinar un período de tiempo durante el que el número de usuarios ha de ser estimado, dividir el período de tiempo en intervalos, proporcionar una pluralidad de matrices de bits correspondientes al número de intervalos, asociar cada matriz de bits de la pluralidad de matrices de bits con un intervalo, de tal manera que por cada intervalo se use una matriz de bits. En cada intervalo, los bits de la serie asociada de bits son activados según los datos dependientes de las entidades que acceden al sitio web durante el intervalo. De esta forma, cada matriz de bits permite estimar el número de usuarios del sitio web que accedió al sitio web durante el intervalo asociado. Después del final del período de tiempo, se determina una matriz de bits que permite estimar el número de usuarios del sitio web que accedió al sitio web durante el período de tiempo. Esta matriz de bits puede ser determinada combinando las matrices de bits de cada intervalo usando una operación lógica O entre los bits de las matrices de bits que tienen el mismo significado.

45 Este método permite almacenar eficientemente datos acerca de entidades que han accedido a la página web durante cada intervalo y combinar eficientemente los datos almacenados por cada intervalo realizando una operación lógica O de muestras.

50 En una realización, la función de mapeado mapea los datos dependientes de las entidades que acceden al sitio web sobre un número del conjunto de números o sobre un bit de la matriz de bits en base a una distribución uniforme. En una realización alternativa, se emplea una función de mapeado no lineal, de tal manera que, por ejemplo, el número de elementos de datos dependientes de las entidades que es mapeado sobre un bit varíe dependiendo de la magnitud del elemento de datos. Esto hace que el método sea más escalable en comparación con el uso de una distribución uniforme. En una realización de la distribución no uniforme, la función de mapeado puede estar compuesta de dos pasos: primero, los datos dependientes de la entidad accedente, por ejemplo, una identificación de usuario, son mapeados a un entero h usando una función hash. En la técnica anterior se describen funciones hash que aseguran que los enteros resultantes estén uniformemente distribuidos entre 0 (incluido) y una constante M (excluida). Posteriormente, se puede aplicar una función no lineal al valor hash resultante.

60 En una realización preferida, durante el paso de estimar el número de usuarios del sitio web en base a los bits activados de la matriz de bits se realiza un análisis de probabilidad máxima para determinar el número de entidades accedentes que muy probablemente reproducirá la configuración dada de bits activados de la serie.

65 En una realización preferida, el paso de estimar los números de usuarios del sitio web en base a los bits activados de la matriz de bits puede incluir los pasos de proporcionar una función de estimación relativa a un número de bits activados con un número de usuarios, determinar un número de bits activados contando el número de bits activados

de la matriz de bits, y estimar el número de usuarios del sitio web en base al número de bits activados de la matriz de bits y la función de estimación. En esta realización, la posición de un bit activado es irrelevante, de tal manera que la estimación puede realizarse eficientemente.

5 Preferiblemente, como el número estimado de usuarios se elige un entero para el que el número esperado de bits activados corresponda muy estrechamente al número observado de bits activados. Dado que la realización exacta de este cálculo es cara, un acercamiento preferido es precalcular el número esperado de bits activados para ciertos números de usuarios y usar estos datos para interpolar el número real de usuarios en base al número observado de bits activados.

10 Preferiblemente, los datos dependientes de la entidad son una identificación de la entidad.

El método según la invención puede ser eficiente y puede permitir estimar el número de usuarios con una precisión aceptable para todas las longitudes de intervalo con un algoritmo uniforme y potencialmente sin la necesidad de estimar exactamente parámetros de muestreo.

15 Además, la invención incluye un sistema para estimar un número de usuarios de un sitio web. El sistema puede incluir un medio para determinar que una entidad accede al sitio web, un medio para determinar datos dependientes de la entidad que accede al sitio web, en reacción al acceso del sitio web, un medio para almacenar dichos datos, un medio para comprimir con pérdidas dichos datos almacenados, y un medio para estimar un número de usuarios del sitio web en base a los datos almacenados comprimidos.

El sistema según la invención tiene las mismas ventajas que el método correspondiente según la invención.

25 El sistema según la invención incluye un almacenamiento que almacena una matriz de bits, un medio para mapear repetidas veces los datos dependientes de las entidades que acceden al sitio web sobre un bit de la matriz de bits según una función de mapeado y para activar el bit de la matriz de bits sobre el que se mapearon los datos, y un medio para estimar el número de usuarios del sitio web en base a los bits activados de la matriz de bits.

30 Preferiblemente, el sistema incluye un medio para determinar un período de tiempo para el que el número de usuarios ha de ser estimado, un medio para dividir el período de tiempo en intervalos, un medio para proporcionar una pluralidad de matrices de bits correspondientes al número de intervalos, un medio para asociar cada matriz de bits de la pluralidad de matrices de bits con un intervalo, de tal manera que por cada intervalo se use una matriz de bits, un medio para activar en cada intervalo los bits de la matriz asociada de bits según los datos dependientes de las entidades que acceden al sitio web durante el intervalo, de tal manera que cada matriz de bits permita estimar el número de usuarios del sitio web que accedieron al sitio web durante el intervalo asociado, y un medio para, después del final del período de tiempo, determinar una matriz de bits que permite estimar el número de usuarios del sitio web que accedieron al sitio web durante el período de tiempo combinando las matrices de bits de cada intervalo usando una operación lógica O entre los bits de las matrices de bits que tienen el mismo significado.

40 En una realización, el medio para mapear repetidas veces incluye una función de mapeado que mapea los datos dependientes de la entidad sobre un número del conjunto de números o sobre un bit de la matriz de bits en base a una distribución uniforme. En una realización alternativa, el medio para mapear repetidas veces incluye una función de mapeado que es no lineal, de tal manera que el número de elementos de datos dependientes de las entidades que es mapeado sobre un bit varíe dependiendo de la magnitud del elemento de datos.

45 En el caso de la distribución no lineal, la función de mapeado puede estar compuesta de dos pasos: primero: los datos dependientes de la entidad accedente, por ejemplo, una identificación de usuario, son mapeados a un entero h usando una función hash. En la técnica anterior se describen funciones hash que aseguran que los enteros resultantes estén uniformemente distribuidos entre 0 (incluido) y una constante M (excluida). Posteriormente, se puede aplicar una función no lineal al valor hash resultante.

50 Preferiblemente, el medio para estimar el número de usuarios del sitio web en base a los bits activados de la matriz de bits usa un análisis de probabilidad máxima para determinar el número de entidades accedentes que es muy probable que reproduzca la configuración dada de bits activados de la serie.

55 En una realización preferida, el medio para estimar el número de usuarios del sitio web en base a los bits activados de la matriz de bits incluye un medio de estimación incluyendo una función de estimación relativa a un número de bits activados con un número de usuarios, un medio para determinar un número de bits activados contando el número de bits activados de la matriz de bits, y un medio para estimar el número de usuarios del sitio web en base al número de bits activados de la matriz de bits y la función de estimación.

60 Preferiblemente, el medio para estimar el número de usuarios del sitio web elige como el número estimado de usuarios un entero para el que el número esperado de bits activados corresponde muy estrechamente al número observado de bits activados. Dado que la realización exacta de este cálculo es cara, un acercamiento preferido es precalcular el número esperado de bits activados para ciertos números de usuarios y usar estos datos para

65

interpolar el número real de usuarios en base al número observado de bits activados.

Preferiblemente, los datos dependientes de la entidad son una identificación de la entidad.

5 El método y el sistema según la invención pueden implementarse usando software. Consiguientemente, la invención incluye un producto de programa de ordenador que incluye un medio legible por ordenador y un programa de ordenador grabado en él en forma de una serie de elementos de estado correspondientes a instrucciones que están adaptadas para ser procesadas por un medio de procesado de datos de un aparato de procesado de datos de tal manera que se realice un método según la invención o que se forme un sistema según la invención en el medio de procesado de datos.

Realizaciones preferidas y otros detalles de la presente invención se explicarán a continuación con referencia a las figuras.

15 La figura 1 representa una realización de un sistema para estimar un número de usuarios de un sitio web.

La figura 2 representa un proceso que genera datos de acceso según una realización de la presente invención.

20 La figura 3 representa una realización de un proceso de compresión de datos y estimación de usuario según la invención.

La figura 4 representa un proceso de compresión de datos y estimación de usuario según antecedentes de la invención.

25 La figura 5 representa una realización del método según la invención.

La figura 6 representa una realización del sistema según la presente invención.

30 La figura 1 representa una realización de un sistema según la presente invención. Un servidor web 101 recibe accesos del sitio web 102 y genera un registro de acceso 103 en base a los accesos. El registro de acceso es analizado por un motor de analítica web 104. El motor de analítica web 104 incluye un analizador de registro 105 que analiza el registro de acceso 103 y genera datos de acceso que son almacenados en una base de datos 106. Los datos de acceso 106 son comprimidos regularmente usando un algoritmo de compresión con pérdidas. Un medio para comprimir con pérdidas los datos 107 genera una base de datos 108 para la estimación de un número de usuarios del sitio web. En base a esta base de datos, se estima un número de usuarios del sitio web 109.

40 La figura 2 representa un proceso que genera datos de acceso 106 que pueden ser usados en una realización del método según la presente invención. En el paso 201, se determina que una entidad accede al sitio web. En el paso 202 se determina una identificación de la entidad que accede al sitio web. En el paso 203 se almacenan los datos dependientes de la identificación de la entidad y el proceso se reanuda en el paso 201.

45 La figura 3 representa una realización de un procedimiento de compresión con pérdidas en combinación con un paso de estimación de usuario según la presente invención. En el paso 301, el procedimiento comprueba si se ha llegado al final de un intervalo para el que las entidades accedentes están registradas. Si se ha llegado al final del intervalo, los datos almacenados son comprimidos con pérdidas en el paso 302. El paso 303 comprueba si se ha llegado al final del período de tiempo. En otros términos, se verifica si sigue otro intervalo. Si éste es el caso, el proceso se reanuda en el paso 301. Si se ha llegado al final del período de tiempo, se estima un número de usuarios en el paso 304 en base a los datos almacenados.

50 La figura 4 representa un procedimiento de compresión de datos y estimación de usuario según antecedentes de la invención. Se supone que se está ejecutando un proceso que genera de forma continua datos de acceso análogos, por ejemplo, al proceso representado en la figura 2. En el paso 401, se verifica si se ha superado la cantidad máxima de datos almacenados. Si éste es el caso, el conjunto de números se trunca en el paso 402 y los números almacenados que corresponden a los números truncados del conjunto de números se borran en el paso 403. Posteriormente, se verifica si se ha llegado al final del período de tiempo en el paso 404. Si no se ha llegado al final del período de tiempo, el procedimiento se reanuda en el paso 401. Si se ha llegado al final del período de tiempo, se determina una relación. Como se representa en el paso 405, la relación puede ser, por ejemplo, el tamaño del conjunto de los números truncados más el tamaño del conjunto de los números no truncados, es decir, el tamaño general del conjunto de números al inicio, dividido por el tamaño del conjunto de los números no truncados. En base a esta relación y los datos almacenados, se estima un número de usuarios del sitio web en el paso 406.

65 El ejemplo siguiente ilustra mejor este procedimiento. Se supone que la cantidad máxima de datos a almacenar se pone de manera que sea 5, mientras que el conjunto de números al que se mapean las identificaciones de las entidades son los números entre 0 y 99. Se supone que en un primer intervalo, se almacenan datos que indican que las entidades 1, 12, 17, 19, 33, 43, 67, 82, 83, 92 accedieron a la página web en el intervalo 1. Dado que la cantidad de datos almacenados excede de 5, la mitad inferior del conjunto de números se trunca. Los números 67, 82, 83, y

92 permanecen.

En el intervalo 2, las entidades 1, 3, 72, y 73 acceden a la página web. Los números 72 y 73 son almacenados, de tal manera que las entidades 67, 72, 73, 82, 83, y 92 sean almacenadas. Esto excede de nuevo de la cantidad máxima de datos a almacenar de tal manera que el tercer cuarto del conjunto de números se trunca, es decir, los números de 50 a 74. Los números 82, 83, y 92 permanecen. Éstas son tres entidades. Es conocido que solamente un cuarto de todo el conjunto de números permanece. Por lo tanto, los números de entidades accedentes se multiplica por 4 para estimar el número de usuarios que accedieron a la página web. 3 por 4 con 12. El número exacto de entidades accedentes era 14. Esto quiere decir que el procedimiento descrito proporciona una estimación aceptable del número real de usuarios del sitio web, pero usando solamente una fracción del almacenamiento que sería necesario para almacenar las identidades de todas las entidades accedentes.

En este ejemplo, la cantidad máxima de datos elegida era 5 para proporcionar un ejemplo conciso. En la práctica, valores de 10000 o más darán lugar a una mejor precisión sin incrementar indebidamente la demanda de memoria.

Además, es posible agregar datos de acceso en múltiples escalas al mismo tiempo. Por ejemplo, puede ser necesario preparar una estadística diaria, semanal y anual. En ese caso, las cantidades de truncamiento aplicadas en las diferentes escalas pueden variar. Especialmente los intervalos más cortos se truncarán normalmente menos. Si los datos existentes de intervalos más cortos se combinan con un intervalo mayor, el conjunto de datos combinados debe truncarse al menos tanto como el intervalo más truncado.

La figura 5 representa una realización del método según la presente invención. En el paso 501, se determina que una entidad accede al sitio web. En el paso 502, se determina una identificación de la entidad que accede al sitio web. La identificación de la entidad es mapeada después sobre un bit de una primera matriz de bits según una función de aplicación en el paso 503. El bit de la matriz de bits sobre el que se mapeó la identificación es activado consiguientemente (paso 504).

En una realización preferida se usa la función de mapeado siguiente $H(x)$:

$$H(x) = -K \ln \frac{x+C}{M+C}$$

x denota el código hash derivado de una entidad accedente. Se caracteriza por un número tomado de un intervalo de 0 a $M-1$, por ejemplo, por un valor de 64 bits si $M = 2^{64}$. Por cada intervalo se facilita una matriz de bits que tiene S bits. Cuanto mayor sea S elegido, más alta es la precisión de la estadística generada. Los valores razonables para S están en torno a 1 millón. La función de mapeado $H(x)$ mapea cada entidad sobre un valor de entre 0 y $S-1$. La función de mapeado posee la propiedad de que algunos valores se eligen más a menudo que otros, es decir, la función de mapeado no es lineal. C y K son constantes que pueden derivarse del valor máximo esperado E del parámetro estadístico a examinar y los valores de S y M :

$$C = \frac{M}{E-1}$$

$$K = \frac{S}{\ln E}$$

Con esta función de mapeado para cada entidad se calcula un compartimiento hash $H(x)$ y el bit correspondiente se pone en la matriz de bits. Para combinar las matrices de bits de los intervalos, las matrices de bits pueden combinarse simplemente usando una operación lógica O.

En otra realización preferible de la función de mapeado, C para la función de mapeado $H(x)$ se calcula como:

$$C = \frac{S}{E \cdot \ln\left(\frac{1}{C}\right)}$$

Para determinar C , hay que calcular un punto fijo que se puede estimar por ejemplo con

$$C = \frac{S}{E \cdot \ln\left(\frac{E}{S}\right)}$$

Entonces, K se establece como sigue:

$$K = \frac{S}{\ln\left(\frac{M+C}{C}\right)}$$

5

En esta realización, el parámetro x debe estar uniformemente distribuido. Las técnicas para calcular x a partir de una identificación de una entidad son conocidas, por ejemplo, puede usarse una función hash criptográfica, pero también funciones hash más simples darán normalmente resultados satisfactorios.

10

La función de mapeado descrita exhibe en un amplio rango una calidad invariante de escala de la estimación, si E se elige suficientemente alto. Aunque E sea órdenes de magnitud demasiado altos, la precisión resultante sólo se degradará moderadamente.

15

En el paso 505 representado en la figura 5, se determina si se ha llegado al final del primer intervalo. Si éste no es el caso, el procedimiento se reanuda en el paso 501. En caso contrario, el procedimiento pasa al paso 506, en el que se determina de nuevo que una entidad accede al sitio web. En el paso 507, se determina una identificación de la entidad accedente, y en el paso 508, esta identificación de la entidad es mapeada sobre un bit de una segunda matriz de bits según la función de mapeado antes descrita. El bit correspondiente de la segunda matriz de bits es activado consiguientemente en el paso 509. En el paso 510, se determina si se ha llegado al final del segundo intervalo. Si todavía no se ha llegado al final, el procedimiento se reanuda en el paso 506. De otro modo, el procedimiento pasa al paso 511, en el que la primera matriz de bits y la segunda matriz de bits se combinan usando una operación lógica O, donde el resultado se escribe en la primera matriz de bits. Esto quiere decir que, después del paso 511, la primera matriz de bits indica las identificaciones de las entidades que han accedido al sitio web durante el primer y el segundo intervalo.

20

25

Posteriormente, se verifica si se ha llegado al final del período de tiempo en el paso 512. Si hay más intervalos, el procedimiento pasa al paso 513, donde se borra la segunda matriz de bits. Entonces, el procedimiento salta de nuevo al paso 506. Si se ha llegado al final del período de tiempo en el paso 512, el procedimiento pasa al paso 514 en el que el número de los usuarios del sitio web se estima en base a la primera matriz de bits.

30

La figura 6 representa una realización de un sistema según la invención. El sistema para estimar un número de usuarios de un sitio web 601 incluye un medio para determinar que una entidad 602 accede al sitio web, un medio para determinar una identificación de la entidad que accede al sitio web 603, un medio para almacenar datos dependientes de la identificación de la entidad 604, un medio para comprimir con pérdidas los datos almacenados 605, y un medio para estimar un número de usuarios del sitio web en base a los datos almacenados 606.

35

Las especificaciones y los dibujos se han de considerar ilustrativos más bien que restrictivos. Es evidente que se puede hacer varias modificaciones y cambios en ellos, sin apartarse del alcance de la invención expuesto en las reivindicaciones. Es posible combinar las características descritas en las realizaciones en una forma modificada para proporcionar realizaciones adicionales que se optimicen para un cierto escenario de uso. A condición de que tales modificaciones sean fácilmente evidentes a los expertos en la técnica, dichas modificaciones se considerarán implícitamente descritas en las realizaciones antes descritas.

40

45

REIVINDICACIONES

1. Método para estimar un número de entidades accedentes que están accediendo a un sitio web, incluyendo los pasos de:

5 - proporcionar una primera matriz de bits y una segunda matriz de bits, teniendo cada matriz de bits un número de S bits, donde la primera matriz de bits está asociada con un primer intervalo de tiempo y la segunda matriz de bits está asociada con un segundo intervalo de tiempo de tal manera que durante el primer intervalo de tiempo se use la primera matriz de bits y durante el segundo intervalo de tiempo se use la segunda matriz de bits;

10 - determinar repetidas veces (501, 506) que al sitio web accede una entidad accedente durante los intervalos de tiempo primero y segundo, donde el método incluye en reacción a cada acceso al sitio web por una entidad accedente:

15 - determinar (502) datos de identificación incluyendo un número caracterizante x dependiente de la entidad accedente, **caracterizándose** cada entidad accedente por un número caracterizante x tomado del intervalo de 0 a $M-1$,

20 - calcular un compartimiento hash $H(x)$ para la entidad accedente mapeando (503, 508) el número caracterizante x dependiente de la entidad accedente sobre un valor entre 0 y $S-1$ según una función de mapeado no lineal $H(x)$, y

25 - activar (504) el bit correspondiente al compartimiento hash determinado $H(x)$ para la entidad accedente en la primera matriz de bits cuando el acceso al sitio web se determina durante el primer intervalo de tiempo y activar (509) el bit correspondiente al compartimiento hash determinado $H(x)$ para la entidad accedente en la segunda matriz de bits cuando el acceso al sitio web se determina durante el segundo intervalo de tiempo,

30 donde la primera matriz de bits permite estimar el número de entidades accedentes al sitio web que accedieron al sitio web durante el primer intervalo de tiempo y la segunda matriz de bits permite estimar el número de entidades accedentes al sitio web que accedieron al sitio web durante el segundo intervalo de tiempo;

- combinar (511) las matrices de bits primera y segunda usando una operación lógica O entre los bits de las matrices de bits que tienen el mismo significado, escribiéndose el resultado en la primera matriz de bits;

35 - borrar (513) la segunda matriz de bits; y

- estimar (514) el número de entidades accedentes al sitio web que accedieron al sitio web durante los intervalos de tiempo primero y segundo en base a los bits activados de la primera matriz de bits.

40 2. Método según la reivindicación 1, **caracterizado porque** el paso de estimar el número de entidades accedentes del sitio web en base a los bits activados de la primera matriz de bits usa un análisis de probabilidad máxima que se basa en la configuración de los bits activados de la primera matriz de bits.

45 3. Método según la reivindicación 1, **caracterizado porque** el paso de estimar el número de entidades accedentes del sitio web en base a los bits activados de la primera matriz de bits incluye los pasos de:

- proporcionar una función de estimación que relaciona un número de bits activados de la primera matriz de bits con un número de entidades accedentes,

50 - determinar un número de bits activados contando el número de bits activados de la primera matriz de bits, y

- estimar el número de usuarios del sitio web en base al número de bits activados de la primera matriz de bits según la relación dada por la función de estimación.

55 4. Sistema para estimar un número de entidades accedentes que están accediendo a un sitio web, incluyendo:

60 - un medio de almacenamiento (604) para almacenar una primera matriz de bits y una segunda matriz de bits, teniendo cada matriz de bits un número de S bits, donde la primera matriz de bits está asociada con un primer intervalo de tiempo y la segunda matriz de bits está asociada con un segundo intervalo de tiempo de tal manera que durante el primer intervalo de tiempo se use la primera matriz de bits y durante el segundo intervalo de tiempo se use la segunda matriz de bits;

- un medio (602) para determinar repetidas veces que al sitio web accede una entidad accedente durante los intervalos de tiempo primero y segundo, incluyendo:

65 - un medio (603) para determinar, en reacción a cada acceso del sitio web por una entidad accedente, datos de identificación incluyendo un número caracterizante x dependiente de la entidad accedente, **caracterizándose** cada

entidad accedente por un número caracterizante x tomado del intervalo de 0 a $M-1$,

5 - un medio para calcular un compartimiento hash $H(x)$ para la entidad accedente mapeando el número caracterizante x dependiente de la entidad accedente sobre un valor de entre 0 y $S-1$ según una función de mapeado no lineal $H(x)$, y

10 - un medio para activar el bit correspondiente al compartimiento hash determinado $H(x)$ para la entidad accedente en la primera matriz de bits cuando el acceso del sitio web se determina durante el primer intervalo de tiempo y activando el bit correspondiente al compartimiento hash determinado $H(x)$ para la entidad accedente en la segunda matriz de bits cuando el acceso del sitio web se determina durante el segundo intervalo de tiempo,

15 donde la primera matriz de bits permite estimar el número de entidades accedentes al sitio web que accedieron al sitio web durante el primer intervalo de tiempo y la segunda matriz de bits permite estimar el número de entidades accedentes al sitio web que accedieron al sitio web durante el segundo intervalo de tiempo;

- un medio (605) para combinar las matrices de bits primera y segunda usando una operación lógica O entre los bits de las matrices de bits que tienen el mismo significado, escribiéndose el resultado en la primera matriz de bits;

20 - un medio para borrar la segunda matriz de bits; y

- un medio (606) para estimar el número de entidades accedentes al sitio web que accedieron al sitio web durante los intervalos de tiempo primero y segundo en base a los bits activados de la primera matriz de bits.

25 5. Sistema según la reivindicación 4, **caracterizado porque**

el medio para estimar el número de entidades accedentes del sitio web en base a los bits activados de la primera matriz de bits usa un análisis de probabilidad máxima que se basa en la configuración de los bits activados de la primera matriz de bits.

30 6. Sistema según la reivindicación 4 o 5, **caracterizado porque** el medio para estimar el número de entidades accedentes del sitio web en base a los bits activados de la primera matriz de bits incluye:

35 - un medio de estimación incluyendo una función de estimación que relaciona un número de bits activados de la primera matriz de bits con un número de entidades accedentes,

- un medio para determinar un número de bits activados contando el número de bits activados de la primera matriz de bits, y

40 - un medio para estimar el número de entidades accedentes del sitio web en base al número de bits activados de la primera matriz de bits según la relación dada por la función de estimación.

45 7. Un producto de programa de ordenador, incluyendo el producto de programa de ordenador un medio legible por ordenador y un programa de ordenador grabado en él en forma de una serie de elementos de estado correspondientes a instrucciones que están adaptadas para ser procesadas por un medio de procesamiento de datos de un aparato de procesamiento de datos de tal manera que se realice un método según al menos una de las reivindicaciones 1 a 3 o se forme un sistema según al menos una de las reivindicaciones 4 a 6 en el medio de procesamiento de datos.

FIG 1

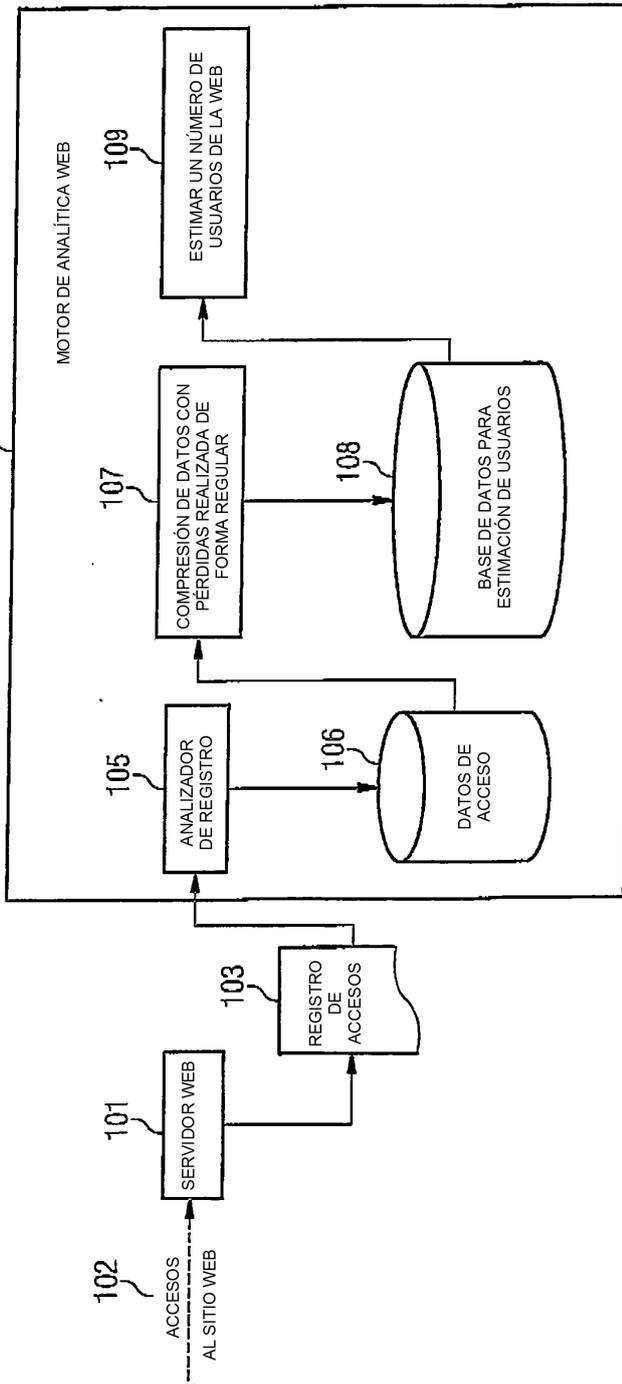


FIG 2

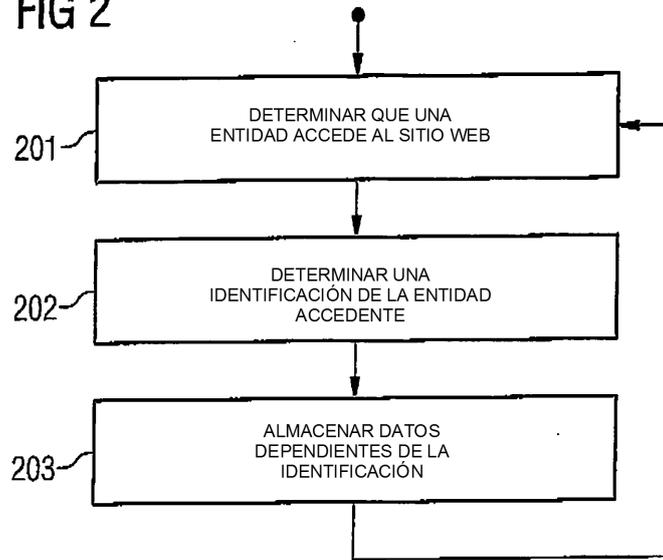


FIG 3

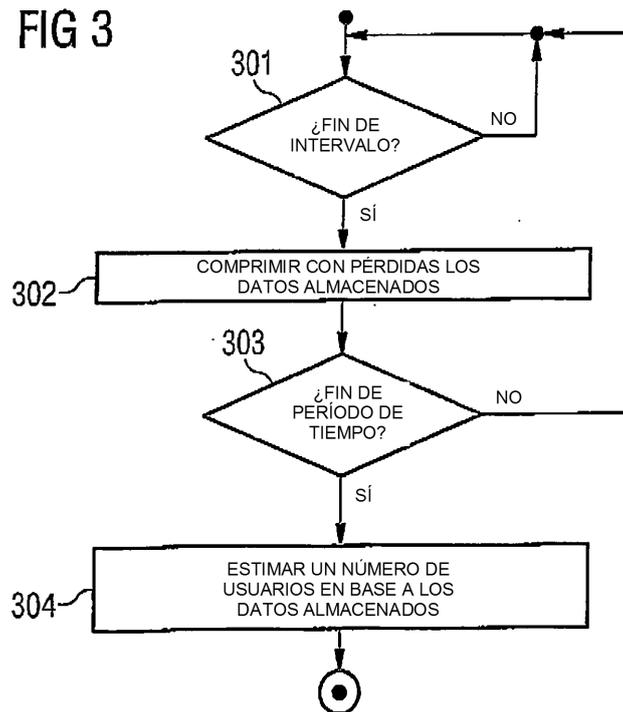


FIG 4

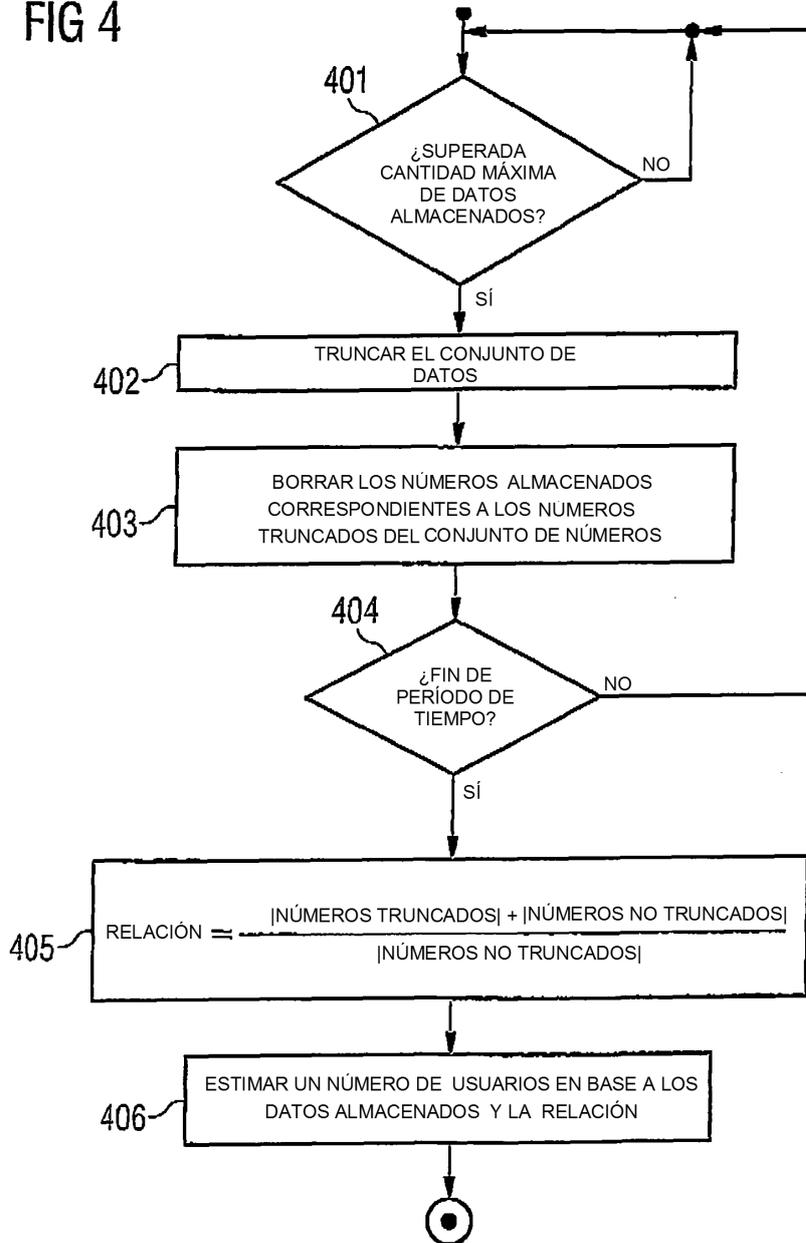


FIG 5

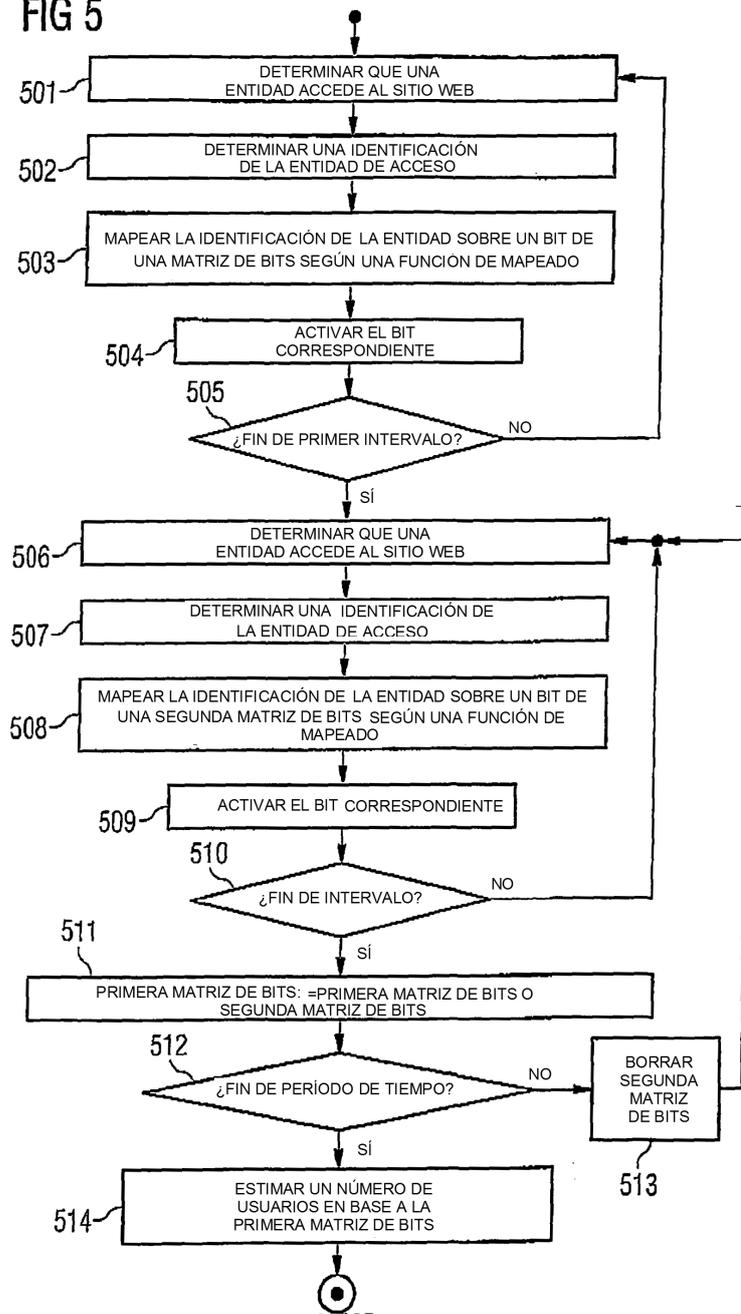


FIG 6

