

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 642 379**

51 Int. Cl.:

G06F 17/30 (2006.01)

G06F 17/27 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.11.2011 E 13175515 (9)**

97 Fecha y número de publicación de la concesión europea: **30.08.2017 EP 2650805**

54 Título: **Dispositivo de registro de palabras relacionadas, dispositivo de procesamiento de información, método de registro de palabras relacionadas, programa para dispositivo de registro de palabras relacionadas, y medio de almacenamiento**

30 Prioridad:

10.11.2010 JP 2010252326

10.11.2010 JP 2010252325

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

16.11.2017

73 Titular/es:

RAKUTEN, INC. (100.0%)

1-14-1, Tamagawa

Setagaya-ku, Tokyo 158-0094, JP

72 Inventor/es:

HIRATE, YU

74 Agente/Representante:

FÚSTER OLAGUIBEL, Gustavo Nicolás

ES 2 642 379 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

Dispositivo de registro de palabras relacionadas, dispositivo de procesamiento de información, método de registro de palabras relacionadas, programa para dispositivo de registro de palabras relacionadas, y medio de almacenamiento

5

CAMPO TÉCNICO

10

La presente invención se refiere a un campo técnico de un dispositivo de registro de palabras relacionadas, un dispositivo de procesamiento de información, un método de registro de palabras relacionadas, un programa para un dispositivo de registro de palabras relacionadas, y un medio de almacenamiento para el registro de una palabra clave de búsqueda como una palabra relacionada.

ANTECEDENTES DE LA TÉCNICA

15

20

Se usa una base de datos de palabras tales como un diccionario de sinónimos para aumentar la precisión del complemento de una palabra clave de búsqueda, una traducción, y similares, en un sistema de búsqueda de información, un sistema de procesamiento de lenguaje natural, un sistema de traducción a máquina, y similares. Se están estudiando diversos dispositivos para constituir automáticamente un diccionario de sinónimos. Por ejemplo, el documento de patente 1 describe un dispositivo de registro de diccionario semántico para proporcionar automáticamente información semántica a una entrada del diccionario de entrada mediante el uso de la información de una entrada de diccionario al que ya se ha proporcionado información semántica en un diccionario semántico.

LISTA DE REFERENCIAS

25

DOCUMENTO DE PATENTE

Documento de Patente 1: JP 2000-268035 A

30

El documento US 6,169,986 B1 describe un sistema y un procedimiento para afinar consultas de búsqueda. El motor de búsqueda sugiere términos relacionados al usuario para permitir que el usuario afine una búsqueda. Los términos relacionados se generan usando datos de correlación de términos de consulta que reflejan la frecuencia con la que han aparecido previamente términos específicos en la misma consulta. Los datos de correlación se generan y se almacenan en una tabla de consulta usando un proceso fuera de línea que analiza un archivo de registro de consultas.

35

RESUMEN DE LA INVENCION**PROBLEMA A RESOLVER POR LA INVENCION**

40

En la técnica divulgada en el documento de patente 1, sin embargo, como se determina un sinónimo basándose en la distancia semántica entre palabras simples de datos de entrada y palabras de todos los datos registrados, la precisión de la determinación de un sinónimo es baja, y el registro de una palabra relacionada, tal como un sinónimo innecesario, no puede suprimirse. Por tanto, es difícil construir un diccionario de palabras relacionadas con alta precisión, tales como un diccionario de sinónimos.

45

La presente invención se ha desarrollado en vista de tal problema, y un ejemplo de un objeto es proporcionar un dispositivo de registro de palabras relacionadas y similares capaz de registrar una palabra relacionada de alta precisión.

50

MEDIOS PARA RESOLVER EL PROBLEMA

55

60

65

El problema anterior se resuelve mediante la materia objeto de las reivindicaciones independientes. Las reivindicaciones dependientes describen modos de realización ventajosos. Un aspecto de implementación a modo de ejemplo incluye: un medio de generación de grupos candidatos de palabras relacionadas que genera un grupo candidato de palabras relacionadas mediante la extracción de candidatos de una palabra relacionada en base a una condición predeterminada de un registro de consulta de búsqueda en el que se almacenan los registros relativos a una consulta de búsqueda; un medio de recepción que recibe una consulta de búsqueda de una palabra de búsqueda introducida por el usuario; un medio de generación de cadenas de caracteres parciales que genera una cadena de caracteres parcial a partir de una cadena de caracteres de la palabra de búsqueda; un medio de extracción de cadenas de caracteres candidatas que extrae una cadena de caracteres candidata del grupo candidato de palabras relacionadas en base a la cadena de caracteres generada parcial; un medio de cálculo de puntuación que calcula una puntuación de idoneidad de la cadena de caracteres candidata en base a la similitud entre la cadena de caracteres candidata y la palabra de búsqueda, el número de veces de uso de la cadena de caracteres candidata, y el número de resultados de búsqueda de la búsqueda de la cadena de caracteres candidata; un medio de clasificación que clasifica las cadenas de caracteres candidatas en orden según las puntuaciones; un medio de generación de criterios que genera una línea de referencia de una puntuación de idoneidad para la clasificación

- 5 como criterio para determinar si la cadena de caracteres candidata está registrada como una palabra relacionada o no en base a la puntuación de idoneidad y la clasificación de la cadena de caracteres candidata; un medio de extracción de cadenas de caracteres de registro que extrae, como una cadena de caracteres de registro a registrar como una palabra relacionada, una cadena de caracteres candidata cuya puntuación de idoneidad está separada de la línea de referencia por un umbral predeterminado o mayor; y un medio de registro de palabras relacionadas que registra la cadena de caracteres de registro extraída y la palabra de búsqueda como palabras relacionadas en una base de datos de palabras relacionadas.
- 10 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de generación de cadenas de caracteres parciales establece la longitud de una cadena de caracteres parciales en base al número de caracteres de la palabra de búsqueda y genera una cadena de caracteres parciales que tiene la longitud fijada.
- 15 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de generación de cadenas de caracteres parciales establece una anchura de desplazamiento del número de caracteres de una palabra de búsqueda en el momento de generar una cadena de caracteres parcial en base al número de caracteres de la palabra de búsqueda y genera una cadena de caracteres parcial con el ancho de desplazamiento fijado.
- 20 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de generación de cadenas de caracteres parciales genera la cadena de caracteres parcial para una palabra de búsqueda obtenida por fonograma que convierte la palabra de búsqueda.
- 25 Otro aspecto está caracterizado por que el dispositivo de registro de palabras relacionadas incluye, además, un medio de ajuste de parámetro que ajusta la información sobre la similitud y el número de veces de uso del medio de cálculo de puntuación.
- 30 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de generación de criterios obtiene la línea de referencia mediante una función aproximada que aproxima la relación entre la clasificación y la puntuación o datos de idoneidad de la clasificación en un intervalo predeterminado.
- 35 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de extracción de cadenas de caracteres de registro establece un límite superior al número de cadenas de caracteres de registro extraídas.
- Otro aspecto está caracterizado por que el dispositivo de registro de palabras relacionadas incluye, además, un medio de almacenamiento de palabras de búsqueda que almacena una palabra de búsqueda relacionada con un producto que se vende en un sitio EC interior o exterior.
- 40 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de cálculo de puntuación calcula la puntuación de idoneidad en base a los datos de un registro en la consulta de búsqueda.
- 45 Otro aspecto está caracterizado por que el dispositivo de registro de palabras relacionadas incluye adicionalmente: un medio de almacenamiento de consultas de búsqueda que almacena las consultas de búsqueda recibidas según el orden de recepción; un medio de extracción de consultas de búsqueda que extrae, del medio de almacenamiento de consultas de búsqueda, una consulta de búsqueda anterior cuyo orden de recepción es anterior al de la consulta de búsqueda recibida en base a una condición de extracción de consultas de búsqueda predeterminada; un medio de almacenamiento de grupos de cadenas de caracteres que almacena, como un conjunto de cadenas de caracteres, una palabra de búsqueda anterior que construye la consulta de búsqueda anterior extraída y una palabra de búsqueda que construye la consulta de búsqueda recibida; un medio de extracción de cadenas de caracteres que extrae un conjunto de cadenas de caracteres que tiene la palabra de búsqueda que es la misma o similar a la palabra de búsqueda anterior del medio de almacenamiento de conjuntos de cadenas de caracteres de acuerdo con una condición de inicio de extracción de conjuntos de cadenas de caracteres predeterminada; y un medio de especificación de palabras relacionadas que especifica un conjunto de caracteres como una palabra relacionada del conjunto de cadenas de caracteres extraído en base a una condición de registro predeterminada, donde el medio de registro de palabras relacionadas registra el conjunto de cadenas de caracteres especificado como palabras relacionadas en una base de datos de palabras relacionadas.
- 50
- 55
- 60 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de almacenamiento de consulta de búsqueda almacena adicionalmente información de identificación del usuario, y el medio de extracción de consulta de búsqueda extrae la consulta de búsqueda anterior cuya información de identificación de usuario coincide con la del usuario que introdujo la consulta de búsqueda como la condición de extracción de consulta de búsqueda.
- 65 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de extracción de consultas de búsqueda extrae una consulta de búsqueda anterior recibida en un tiempo

predeterminado desde el momento de recepción de la consulta de búsqueda como la condición de extracción de consulta de búsqueda.

5 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de almacenamiento de consulta de búsqueda almacena adicionalmente información de categorías, y el medio de extracción de consulta de búsqueda extrae la consulta de búsqueda anterior en base a la información de categorías como la condición de extracción de consulta de búsqueda.

10 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de extracción de cadenas de caracteres extrae el conjunto de cadenas de caracteres en el caso en el que el número de consultas de búsqueda en el medio de almacenamiento de consultas de búsqueda o el número de elementos del grupo de conjunto de cadenas de caracteres exceda un umbral predeterminado como la condición de inicio de extracción del conjunto de cadenas de caracteres.

15 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, el medio de extracción de cadenas de caracteres extrae el conjunto de cadenas de caracteres en el caso en el que transcurre el tiempo predeterminado desde que se satisface la condición de inicio de extracción del conjunto de cadenas de caracteres.

20 Otro aspecto está caracterizado por que, en el dispositivo de registro de palabras relacionadas, en el caso en el que el número de conjuntos de cadenas de caracteres que tienen la misma o similar palabra de búsqueda o la relación de los conjuntos de cadenas de caracteres que tienen la misma o similar palabra de búsqueda sobrepasa un umbral predeterminado en los conjuntos de cadenas de caracteres que tienen la misma o similar palabra de búsqueda anterior como la condición de registro, el medio de especificación de palabras relacionadas especifica el conjunto de cadenas de caracteres como palabras relacionadas.

30 Otro aspecto se refiere a un dispositivo de procesamiento de información para procesar información con referencia a una base de datos de palabras relacionadas del dispositivo de registro de palabras relacionadas, que incluye: un medio de extracción de palabras relacionadas que extrae una palabra relacionada correspondiente a una palabra de búsqueda de la consulta de búsqueda recibida con referencia a la base de datos de palabras relacionadas; y un medio de salida de palabras relacionadas que emite la palabra relacionada extraída por el medio de extracción de palabras relacionadas.

35 Otro aspecto se refiere a un método de registro de palabras relacionadas de un dispositivo de registro de palabras relacionadas para registrar una palabra relacionada, que incluye: una etapa de generación de grupos candidatos de palabras relacionadas mediante la extracción de candidatos de una palabra relacionada en base a una condición predeterminada de un registro de consulta de búsqueda en el que se almacenan los registros con respecto a una consulta de búsqueda; una etapa de recepción para recibir una consulta de búsqueda de una palabra de búsqueda introducida por el usuario; una etapa de generación de cadenas de caracteres parcial para generar una cadena de caracteres parcial a partir de una cadena de caracteres de la palabra de búsqueda; una etapa de extracción de cadenas de caracteres candidatas para extraer una cadena de caracteres candidata del grupo candidato de palabras relacionadas en base a la cadena de caracteres parcial generada; una etapa de cálculo de puntuación para calcular una puntuación de idoneidad de la cadena de caracteres candidata en base a la similitud entre la cadena de caracteres candidata y la palabra de búsqueda, el número de veces de uso de la cadena de caracteres candidata, y el número de resultados de búsqueda de la búsqueda de la cadena de caracteres candidata; una etapa de clasificación para clasificar las cadenas de caracteres candidatas según el orden de puntuación; una etapa de generación de criterio para generar una línea de referencia de una puntuación de idoneidad para la clasificación como un criterio para determinar si la cadena de caracteres candidata está registrada como una palabra relacionada o no en base a la puntuación de idoneidad y la clasificación de la cadena de caracteres candidata; una etapa de extracción de cadenas de caracteres de registro para extraer, como una cadena de caracteres de registro a registrar como una palabra relacionada, una cadena de caracteres candidata cuya puntuación de idoneidad está separada de la línea de referencia por un umbral predeterminado o mayor; y una etapa de registro de palabras relacionadas para registrar la cadena de caracteres de registro extraída y la palabra de búsqueda como palabras relacionadas en una base de datos de palabras relacionadas.

55 Otro aspecto se refiere a un ordenador que funciona como: un medio de generación de grupos candidatos de palabras relacionadas que genera un grupo candidato de palabras relacionadas mediante la extracción de candidatos de una palabra relacionada en base a una condición predeterminada de un registro de consulta de búsqueda en el que se almacenan los registros con respecto a una consulta de búsqueda; un medio de recepción que recibe una consulta de búsqueda de una palabra de búsqueda introducida por el usuario; un medio de generación de cadenas de caracteres parciales que genera una cadena de caracteres parcial a partir de una cadena de caracteres de la palabra de búsqueda; un medio de extracción de cadenas de caracteres candidatas que extrae una cadena de caracteres candidata del grupo candidato de palabras relacionadas en base a la cadena de caracteres generada parcial; un medio de cálculo de puntuación que calcula una puntuación de idoneidad de la cadena de caracteres candidata en base a la similitud entre la cadena de caracteres candidata y la palabra de búsqueda, el número de veces de uso de la cadena de caracteres candidata, y el número de resultados de

búsqueda de la búsqueda de la cadena de caracteres candidata; un medio de clasificación que clasifica las cadenas de caracteres candidatas en orden de las puntuaciones; un medio de generación de criterio que genera una línea de referencia de una puntuación de idoneidad para la clasificación como criterio para determinar si la cadena de caracteres candidata está registrada como una palabra relacionada o no en base a la puntuación de idoneidad y la clasificación de la cadena de caracteres candidata; un medio de extracción de cadenas de caracteres de registro que extrae, como una cadena de caracteres de registro a registrar como una palabra relacionada, una cadena de caracteres candidata cuya puntuación de idoneidad está separada de la línea de referencia por un umbral predeterminado o mayor; y un medio de registro de palabras relacionadas que registra la cadena de caracteres de registro extraída y la palabra de búsqueda como palabras relacionadas en una base de datos de palabras relacionadas.

Otro aspecto se refiere a almacenar un programa para un dispositivo de registro de palabras relacionadas, para hacer que un ordenador funcione como: un medio de generación de grupos candidatos de palabras relacionadas que genera un grupo candidato de palabras relacionadas mediante la extracción de candidatos de una palabra relacionada en base a una condición predeterminada de un registro de consulta de búsqueda en el que se almacenan los registros con respecto a una consulta de búsqueda; un medio de recepción que recibe una consulta de búsqueda de una palabra de búsqueda introducida por el usuario; un medio de generación de cadenas de caracteres parciales que genera una cadena de caracteres parcial a partir de una cadena de caracteres de la palabra de búsqueda; un medio de extracción de cadenas de caracteres candidatas que extrae una cadena de caracteres candidata del grupo candidato de palabras relacionadas en base a la cadena de caracteres generada parcial; un medio de cálculo de puntuación que calcula una puntuación de idoneidad de la cadena de caracteres candidata en base a la similitud a partir de un registro de consulta de búsqueda en el que se almacenan los registros con respecto a una consulta de búsqueda, se recibe una consulta de búsqueda de una palabra de búsqueda introducida por el usuario, se genera una cadena de caracteres parcial a partir de una cadena de caracteres de la palabra de búsqueda, se extrae una cadena de caracteres candidata del grupo candidato de palabras relacionadas en base a las cadenas de caracteres parciales, se calcula una puntuación de idoneidad de la cadena de caracteres candidata en base a la similitud entre la cadena de caracteres candidata y la palabra de búsqueda, el número de veces de uso de la cadena de caracteres candidata, y el número de resultados de búsqueda de la búsqueda de la cadena de caracteres candidata, se clasifican las cadenas de caracteres candidatas en orden de puntuación, se genera una línea de referencia de una puntuación de idoneidad para la clasificación como un criterio para determinar si la cadena de caracteres candidata está registrada como una palabra relacionada o no en base a la puntuación de idoneidad y la clasificación de la cadena de caracteres candidata, se extrae una cadena de caracteres candidata cuya puntuación de idoneidad está separada de la línea de referencia por un umbral predeterminado o mayor como una cadena de caracteres de registro a registrar como una palabra relacionada, y se registran la cadena de caracteres de registro extraída y la palabra de búsqueda como palabras relacionadas en la base de datos de palabras relacionadas. En consecuencia, como se extrae una cadena de caracteres candidata del grupo candidato de palabras relacionadas en base a una cadena de caracteres parcial que tiene un menor número de caracteres, se puede suprimir el registro de una palabra relacionada innecesaria por la puntuación de idoneidad de la cadena de caracteres candidata. Por lo tanto, puede mejorarse la precisión de una palabra relacionada.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La figura 1 es un diagrama esquemático que ilustra un ejemplo de una configuración esquemática de un sistema de registro de palabras relacionadas como un modo de realización de la presente invención.

La figura 2 es un diagrama de bloques que ilustra un ejemplo de una configuración esquemática de un servidor de registro de palabras relacionadas en la figura 1.

La figura 3 es un diagrama de bloques que ilustra un ejemplo de una configuración esquemática de un servidor de suministro de información en la figura 1.

La figura 4 es un diagrama de bloques que ilustra un ejemplo de una configuración esquemática de un terminal en la figura 1.

La figura 5 es un diagrama de flujo que ilustra un ejemplo de las operaciones de generación de un grupo candidato de palabras relacionadas de un servidor de registro de palabras relacionadas en la operación de una implementación a modo de ejemplo de un sistema 1 de registro de palabras relacionadas en la figura 1.

La figura 6 es un diagrama esquemático que ilustra un ejemplo de una página web que se muestra en el terminal en la figura 1.

La figura 7 es un diagrama de flujo que ilustra un ejemplo de las operaciones de registro de una palabra relacionada en una implementación a modo de ejemplo del servidor de registro de palabras relacionadas en la figura 1.

La figura 8 es un diagrama esquemático que ilustra un ejemplo de extracción de cadenas de caracteres parciales.

Las figuras 9A y 9B son diagramas esquemáticos que ilustran un ejemplo de un gráfico de puntuación de clasificación.

5 La figura 10 es un diagrama de flujo que ilustra una subrutina del ejemplo de operación de cálculo de una puntuación de idoneidad en el servidor de registro de palabras relacionadas en la figura 1.

Las figuras 11A a 11C son diagramas que ilustran un ejemplo de un patrón de disponibilidad para el cálculo de la puntuación de idoneidad.

10 La figura 12 es un diagrama esquemático que ilustra un ejemplo de una página web visualizada en un terminal en una operación de un modo de realización del sistema 1 de registro de palabras relacionadas en la figura 1.

15 La figura 13 es un diagrama de flujo que ilustra un ejemplo de operaciones de construcción de una base de datos de registro de consultas de búsqueda del servidor de registro de palabras relacionadas en la figura 1.

La figura 14 es un diagrama esquemático que ilustra un ejemplo de una página web que se muestra en el terminal en la figura 1.

20 La figura 15 es un diagrama de flujo que ilustra un ejemplo de la operación de registro de una palabra relacionada, del servidor de registro de palabras relacionadas en la figura 1.

La figura 16 es un diagrama esquemático que ilustra un ejemplo de un conjunto de cadenas de caracteres.

25 MODOS DE REALIZAR LA INVENCION

En lo sucesivo en el presente documento, se describirán modos de realización de la presente invención con referencia a los dibujos. Los modos de realización que se describirán se refieren al caso de aplicar la presente invención a un sistema de registro de palabras relacionadas.

30 [1. Resumen de configuraciones y funciones del sistema de registro de palabras relacionadas]

En primer lugar, se describirán usando la figura 1 la configuración y las funciones esquemáticas de un sistema de registro de palabras relacionadas de acuerdo con un modo de realización de la presente invención.

35 La figura 1 es un diagrama esquemático que ilustra un ejemplo de una configuración esquemática de un sistema 1 de registro de palabras relacionadas como un modo de realización.

40 Como se ilustra en la figura 1, el sistema 1 de registro de palabras relacionadas incluye: un servidor 10 de registro de palabras relacionadas (un ejemplo de un dispositivo de registro de palabras relacionadas) que registra una palabra relacionada; un servidor 20 de suministro de información instalado para ejecutar un sitio de compras (un ejemplo de un sitio EC (Comercio Electrónico, Electronic Commerce) interior) como un ejemplo de un sitio de suministro de información utilizado por el usuario y para proporcionar información registrada en el servidor 10 de registro de palabras relacionadas e información de productos al usuario; y un terminal 30 utilizado por el usuario para recuperar un producto y similares en el servidor 20 de suministro de información.

45 El servidor 10 de registro de palabras relacionadas y el servidor 20 de suministro de información están conectados entre sí a través de una red de área local o similar, de manera que pueden transmitirse/recibirse datos, y constituyen un sistema 5 de servidor. El sistema 5 de servidor y el terminal 30 están conectados entre sí a través de una red 3 y pueden transmitir/recibir datos a/desde el uno al otro por un protocolo de comunicación (por ejemplo, TCP/IP).
50 Además, la red 3 se construye mediante, por ejemplo, Internet, líneas de comunicación dedicadas (por ejemplo, una línea CATV (Televisión de Antena Comunitaria, Community Antenna Television), una red de comunicaciones móviles (incluyendo una estación de base y similares), una puerta de enlace, y similares.

55 El servidor 10 de registro de palabras relacionadas construye una base de datos de palabras relacionadas a partir de actividades de búsqueda y similares del usuario en el terminal 30 y similares, y proporciona soporte a las actividades de búsqueda del usuario en el terminal 30 y similares.

60 El servidor 20 de suministro de información recibe una consulta de búsqueda que incluye una palabra de búsqueda del terminal 30, realiza una búsqueda, y transmite un resultado de búsqueda al terminal 30. El servidor 20 de suministro de información transmite una consulta de búsqueda recibida al servidor 10 de registro de palabras relacionadas con el fin de construir una base de datos de palabras relacionadas. El servidor 20 de suministro de información realiza, como un sitio de compras, una búsqueda de un producto, suministra información de los productos, suministra información publicitaria, un procedimiento para la compra de un producto, y similares.

65 Existe una pluralidad de terminales 30 de usuario.

[2. Configuraciones y funciones de los servidores]

(2.1 Configuraciones y funciones del servidor 10 de registro de palabras relacionadas)

5 Las configuraciones y funciones del servidor 10 de registro de palabras relacionadas se describirán usando la figura 2.

La figura 2 es un diagrama de bloques que ilustra un ejemplo de una configuración esquemática del servidor 10 de registro de palabras relacionadas.

10 Como se ilustra en la figura 2, el servidor 10 de registro de palabras relacionadas que funciona como un ordenador, tiene una unidad 11 de comunicación, una unidad 12 de almacenamiento, una interfaz 13 de entrada/salida, y una unidad 14 de control del sistema. La unidad 14 de control del sistema y la interfaz 13 de entrada/salida están conectadas entre sí a través de un bus de sistema 15.

15 La unidad 11 de comunicación está conectada a la red 3 para controlar el estado de la comunicación con el terminal 30 o similar y, además, está conectada a la red de área local para transmitir/recibir datos a/desde otro servidor tal como el servidor 20 de suministro de información en la red de área local.

20 La unidad 12 de almacenamiento se construye, por ejemplo, mediante una unidad de disco duro o similar, y almacena diversos programas tales como un sistema operativo y un programa de servidor, datos, y similares. Además, pueden obtenerse los diversos programas, por ejemplo, a partir de otro dispositivo servidor o similares a través de la red 3 o pueden almacenarse en un medio de registro y leerse a través de un dispositivo de disco (no ilustrado).

25 En la unidad 12 de almacenamiento, se construyen una base de datos 12a de registro de consultas de búsqueda (en lo sucesivo en el presente documento, denominada "BD 12a de registro de consultas de búsqueda") que almacena un registro relativo a una consulta de búsqueda recibida desde el terminal 30, una base de datos 12b de candidatos de palabras relacionadas (en lo sucesivo en el presente documento, denominada "BD 12b de candidatos de palabras relacionadas") que almacena, como un grupo candidato de palabras relacionadas, cadenas de caracteres candidatas de una palabra relacionada generada a partir de la consulta de búsqueda, una base de datos 12c de palabras relacionadas (en lo sucesivo en el presente documento, denominada "BD 12c de palabras relacionadas") que almacena una palabra relacionada generada a partir de la consulta de búsqueda, una base de datos 12d de conjuntos de cadenas de caracteres (en lo sucesivo en el presente documento, denominada "BD 12b de conjuntos de cadenas de caracteres"), y similares. Además, la BD 12a de registro de consultas de búsqueda, la BD 12b de candidatos de palabras relacionadas, la BD 12c de palabras relacionadas, y la BD 12b de conjuntos de cadenas de caracteres pueden construirse en otro servidor que no sea el servidor 10 de registro de palabras relacionadas en el sistema 5 de servidor o pueden construirse fuera del sistema 5 de servidor.

40 En la BD 12a de registro de consultas de búsqueda (un ejemplo de un medio de almacenamiento de consultas de búsqueda), se almacena una consulta de búsqueda asociada al tiempo de recepción de una consulta de búsqueda e información de identificación de usuario para discriminar a un usuario tal como un ID de usuario. Además, la información de identificación de usuario incluye un ID de usuario de un sitio de compras o similar, el número de identificación de un terminal de usuario o un punto de acceso, la dirección IP, y similares. Además, cuando el usuario hace una búsqueda de productos en el servidor 20 de suministro de información, se almacena una consulta de búsqueda en la BD 12a de registro de consultas de búsqueda en asociación a la información de categoría, tal como una categoría de productos a la que pertenece el producto o una categoría de producto que se muestra en una página web abierta por el usuario.

50 En la BD 12b de candidatos de palabras relacionadas, se almacenan cadenas de caracteres candidatas de la palabra relacionada extraída en base a una condición predeterminada a partir del registro de consultas de búsqueda en la BD 12a de registro de consultas de búsqueda como un grupo candidato de palabras relacionadas.

55 En la BD 12c de palabras relacionadas, se almacena una palabra relacionada generada por el servidor 10 de registro de palabras relacionadas.

60 En la BD 12d del conjunto de cadenas de caracteres (un ejemplo de un medio de almacenamiento de conjuntos de cadenas de caracteres), se almacena un conjunto de cadenas de caracteres de una palabra de búsqueda en la consulta de búsqueda recibida y una palabra de búsqueda de la consulta de búsqueda extraída de la BD 12a de registro de consultas de búsqueda.

A continuación, la interfaz 13 de entrada/salida realiza un proceso de interfaz entre la unidad 11 de comunicación y la unidad 12 de almacenamiento y la unidad 14 de control del sistema.

65 La unidad 14 de control del sistema está construida por una CPU (Unidad de Procesamiento Central, Central Processing Unit) 14a, una ROM (Memoria de Sólo Lectura, Read Only Memory) 14b, una RAM (Memoria de Acceso

Aleatorio, Random Access Memory) 14c, y similares. La unidad 14 de control del sistema lee y ejecuta diversos programas almacenados en la ROM 14b y la unidad 12 de almacenamiento mediante la CPU 14a para realizar un proceso de registro de palabras relacionadas y similares.

5 (2.2 Configuraciones y funciones del servidor 20 de suministro de información)

A continuación, se describirán usando la figura 3 las configuraciones y funciones del servidor 20 de suministro de información.

10 La figura 3 es un diagrama de bloques que ilustra un ejemplo de una configuración esquemática del servidor 20 de suministro de información.

15 Como se ilustra en la figura 3, el servidor 20 de suministro de información tiene una unidad 21 de comunicación, una unidad 22 de almacenamiento, una interfaz 23 de entrada/salida, y una unidad 24 de control del sistema. La unidad 24 de control del sistema y la interfaz 23 de entrada/salida están conectadas entre sí a través de un bus 25 de sistema. Además, dado que las configuraciones y funciones del servidor 20 de suministro de información son casi las mismas que las del servidor 10 de registro de palabras relacionadas, se describirán principalmente cosas diferentes en las configuraciones y funciones del servidor 10 de registro de palabras relacionadas.

20 La unidad 21 de comunicación controla el estado de comunicación con el terminal 30, el servidor 10 de registro de palabras relacionadas, y similares a través de la red 3, la red de área local, o similar.

25 En la unidad 22 de almacenamiento, se construyen una base de datos 22a de productos (en lo sucesivo en el presente documento, denominada "BD de productos"), una base de datos 22b de miembros (en lo sucesivo en el presente documento, denominada "BD de miembros"), y similares.

30 En la BD 22a de productos, en asociación con un ID de producto como un identificador para identificar un producto, se almacenan el nombre y el tipo de producto, una imagen del producto, especificaciones e información de producto, información publicitaria relacionada con cada uno de los productos, y similares. En la BD 22a de productos, se almacena un archivo de una página web del producto descrita en un lenguaje de marcado, tal como HTML (Lenguaje de Marcado de Hipertexto, Hyper Text Markup Language), un XML (Lenguaje de Marcado Extensible, Extensible Markup Language), o similares.

35 Además, en la BD 22a de productos, se construye una base de datos de búsqueda para recuperar un producto. La BD 22a de productos es un ejemplo de un medio de almacenamiento de palabras de búsqueda que almacena una palabra relacionada con un producto que se vende en un sitio EC en el sistema 5 de servidor.

40 En la BD 22b de miembros, se registra información del usuario, tal como el ID de usuario de un usuario registrado como miembro (usuario de un sitio de compras), nombre, dirección, número de teléfono, correo electrónico, ocupación, aficiones, historial de compras, temas y géneros (categorías de productos) en la que el usuario está interesado, y similares. En la BD 22b de miembros, están registrados el ID de usuario, el ID de inicio de sesión y la contraseña necesaria para que el usuario inicie sesión en el sitio de compras desde el terminal 30. En este caso, el ID de inicio de sesión y la contraseña son información de inicio de sesión usada para el proceso de inicio de sesión (proceso de autenticación de usuario).

45 La unidad 24 de control del sistema incluye una CPU 24a, una ROM 24b y una RAM 24c. La unidad 24 de control del sistema lee y ejecuta diversos programas almacenados en la ROM 24b y la unidad 22 de almacenamiento por la CPU 24a para realizar el proceso de búsqueda de producto, el proceso de compra del producto por el usuario, y similares.

50 (2.3 Configuraciones y funciones del terminal 30)

Las configuraciones y funciones del terminal 30 se describirán usando la figura 4.

55 La figura 4 es un diagrama de bloques que ilustra un ejemplo de la configuración esquemática del terminal 30.

60 Como se ilustra en la figura 4, el terminal 30 que funciona como un ordenador es, por ejemplo, un ordenador personal o un terminal portátil tal como un teléfono inalámbrico portátil, incluyendo un teléfono inteligente o una PDA, y tiene una unidad 31 de comunicación, una unidad 32 de almacenamiento, una unidad 33 de visualización, una unidad 34 de operación, una interfaz 35 de entrada/salida, y una unidad 36 de control del sistema. La unidad 36 de control del sistema y la interfaz 35 de entrada/salida están conectadas entre sí a través de un bus 37 de sistema.

65 La unidad 31 de comunicación controla la comunicación con el servidor 20 de suministro de información o similar a través de la red 3. Además, en el caso en el que el terminal 30 es un dispositivo de terminal portátil, la unidad 31 de comunicación tiene una función de comunicación inalámbrica para la conexión a la red de comunicación móvil en la red 3.

La unidad 32 de almacenamiento es, por ejemplo, una unidad de disco duro o similar, y almacena un sistema operativo, un programa de un navegador web, un programa de una barra de herramientas para el navegador web, y similares.

La 33 unidad de visualización está formada por, por ejemplo, un dispositivo de visualización de cristal líquido, un dispositivo de EL (Electro Luminiscencia, Electro Luminescence), o similares. En la unidad 33 de visualización, se muestra por el navegador web una página web de una pantalla de búsqueda o una página web que muestra el resultado de una búsqueda de producto proporcionada por el servidor 20 de suministro de información.

La unidad 34 de operación consiste, en un teclado, un ratón, y similares. El usuario introduce una respuesta con la unidad 34 de operación. Además, en el caso en el que la unidad 33 de visualización es un panel de visualización de un sistema de conmutador de contacto, tal como una pantalla táctil, la unidad 34 de operación obtiene información de la posición de la unidad 33 de visualización de una posición en la que el usuario toca o a la que se acerca.

La interfaz 35 de entrada/salida es una interfaz entre la unidad 31 de comunicación y la unidad 32 de almacenamiento y la unidad 36 de control del sistema.

La unidad 36 de control del sistema incluye, por ejemplo, una CPU 36a, una ROM 36b y una RAM 36c. La unidad 36 de control del sistema lee y ejecuta diversos programas almacenados en la ROM 36b, RAM 36c y la unidad 32 de almacenamiento por la CPU 36a. Por ejemplo, la unidad 36 de control del sistema ejecuta un programa de un navegador web y funciona como un navegador web.

[3. Funcionamiento de una implementación a modo de ejemplo de un sistema de registro de palabras relacionadas]

A continuación, se describirá usando las figuras 5 a 11 la operación de una implementación a modo de ejemplo del sistema 1 de registro de palabras relacionadas de acuerdo con un modo de realización de la presente invención.

La figura 5 es un diagrama de flujo que ilustra un ejemplo de las operaciones de generación de un grupo candidato de palabras relacionadas en el servidor 10 de registro de palabras relacionadas. La figura 6 es un diagrama esquemático que ilustra un ejemplo de una página web que se muestra en el terminal 30. La figura 7 es un diagrama de flujo que ilustra un ejemplo de operaciones de registro de una palabra relacionada en el servidor de registro de palabras relacionadas. La figura 8 es un diagrama esquemático que ilustra un ejemplo de extracción de una cadena de caracteres parcial. Las figuras 9A y 9B son diagramas esquemáticos que ilustran un ejemplo de un gráfico de puntuación de clasificación. La figura 10 es un diagrama de flujo que ilustra una subrutina del ejemplo de operación de cálculo de una puntuación de idoneidad en el servidor 10 de registro de palabras relacionadas. Las figuras 11A a 11C son diagramas que ilustran un ejemplo de un patrón de disponibilidad para el cálculo de la puntuación de idoneidad.

(3.1 Generación de grupo candidato de palabras relacionadas)

La generación de un grupo de candidatos de palabras relacionadas se describirá usando la figura 5.

En primer lugar, cuando una consulta de búsqueda que incluye una palabra de búsqueda introducida por el usuario se recibe desde el terminal 30, el servidor 20 de suministro de información realiza una búsqueda y transmite la consulta de búsqueda al servidor 10 de registro de palabras relacionadas. Además, el servidor 20 de suministro de información transmite el número de resultados de búsqueda para la consulta de búsqueda al servidor 10 de registro de palabras relacionadas.

El servidor 10 de registro de palabras relacionadas recibe la consulta de búsqueda y el número de resultados de búsqueda del servidor 20 de suministro de información y almacena la consulta de búsqueda junto con el número de resultados de búsqueda en la BD 12a de registro de consultas de búsqueda. El servidor 10 de registro de palabras relacionadas puede almacenar la consulta de búsqueda junto con la hora recibida cuando se recibe la consulta de búsqueda y la información de identificación de usuario para identificar al usuario, tal como el ID de usuario o la dirección IP.

A continuación, como se ilustra en la figura 5, el servidor 10 de registro de palabras relacionadas extrae una consulta de búsqueda del registro de consultas de búsqueda (etapa S1). Específicamente, la unidad 14 de control del sistema en el servidor 10 de registro de palabras relacionadas extrae una consulta de búsqueda de la BD 12a de registro de consultas de búsqueda.

A continuación, el servidor 10 de registro de palabras relacionadas determina si la consulta de búsqueda es una consulta de búsqueda cuyo número de resultados de búsqueda es uno o más (etapa S2). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas lee el número de resultados de búsqueda correspondientes a la consulta de búsqueda extraída de la BD 12a de registro de consultas de búsqueda y determina si el número de resultados de búsqueda es uno o más. Además, la unidad 14 de control del sistema en el

servidor 10 de registro de palabras relacionadas puede transmitir la consulta de búsqueda extraída al servidor 20 de suministro de información para obtener el número de resultados de búsqueda afectados por la consulta de búsqueda y recibir el número de aciertos del servidor 20 de suministro de información.

5 En el caso en el que el número de resultados de búsqueda es uno o más (SÍ, en la etapa S2), el servidor 10 de registro de palabras relacionadas registra la consulta de búsqueda extraída en un grupo candidato de palabras relacionadas (etapa S3). Específicamente, en el caso de una consulta cuyo número de resultados de búsqueda es uno o más, es decir, cuyo número de peticiones de búsqueda no es cero, la unidad 14 de control del sistema en el
10 servidor 10 de registro de palabras relacionadas registra la consulta de búsqueda (incluyendo la palabra de búsqueda como la cadena de caracteres candidata de palabras relacionadas) como un grupo candidato de palabras relacionadas en la BD 12b de candidatos de palabras relacionadas.

15 En el caso en el que el número de resultados de búsqueda no es uno o mayor (NO en la etapa S2), el servidor 10 de registro de palabras relacionadas no registra la consulta de búsqueda extraída en el grupo candidato de palabras relacionadas.

20 A continuación, se determina si queda una consulta de búsqueda o no (etapa S4). Específicamente, la unidad 14 de control del sistema en el servidor 10 de registro de palabras relacionadas determina si una consulta de búsqueda que no se ha sometido a la determinación del número de resultados de búsqueda existe o no en la BD 12a de registro de consultas de búsqueda.

25 En el caso en el que existe una consulta de búsqueda restante (SÍ, en la etapa S4), el servidor 10 de registro de palabras relacionadas regresa a la etapa S1, y extrae una consulta de búsqueda subsiguiente. En el caso en el que no existe una consulta de búsqueda restante (NO en la etapa S4), el proceso de generación del grupo candidato de palabras relacionadas finaliza. De tal manera, el servidor 10 de registro de palabras relacionadas genera de forma preliminar un grupo candidato de palabras relacionadas de una cierta escala. El servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de generación de grupos candidatos de palabras relacionadas que genera un grupo candidato de palabras relacionadas extrayendo candidatos de una palabra relacionada en base
30 a una condición predeterminada a partir de un registro de consultas de búsqueda en el que se almacenan los registros con respecto a una consulta de búsqueda.

35 Además, cuando el número de resultados de búsqueda no es cero en el caso en el que se recibe una consulta de búsqueda y se realiza una búsqueda, el servidor 20 de suministro de información puede añadir información de que el número de resultados de búsqueda no es cero a la consulta de búsqueda y transmitir lo resultante al servidor 10 de registro de palabras relacionadas. En este caso, cuando se recibe la consulta de búsqueda cuyo número de resultados de búsqueda no es cero, el servidor 10 de registro de palabras relacionadas lo almacena en la BD 12b de candidatos de palabras relacionadas.

40 (3.2 Registro de palabra relacionada)

La operación de registro de una palabra relacionada se describirá usando las figuras 6 a 9.

45 En primer lugar, como se ilustra en la figura 6, en la unidad 33 de visualización del terminal 30, una palabra de búsqueda, por ejemplo, "cena de familia A" se introduce por el usuario en un cuadro 41 de entrada de palabra de búsqueda en una página web 40 proporcionada desde el servidor 20 de suministro de información. Se hace clic en un botón 42 de búsqueda para realizar una búsqueda, y el terminal 30 transmite una consulta de búsqueda que incluye la palabra de búsqueda al servidor 20 de suministro de información. A continuación, el servidor 20 de suministro de información recibe la consulta de búsqueda que incluye la palabra de búsqueda introducida por el usuario desde el terminal 30, y la transmite al servidor 10 de registro de palabras relacionadas.
50

55 A continuación, como se ilustra en la figura 7, el servidor 10 de registro de palabras relacionadas recibe la consulta de búsqueda que incluye la palabra de búsqueda introducida por el usuario (etapa S10). Específicamente, la unidad 14 de control del sistema en el servidor 10 de registro de palabras relacionadas recibe la consulta de búsqueda a través de la unidad 11 de comunicación del servidor 20 de suministro de información. Además, como se describe en la generación de un candidato de palabras relacionadas, el servidor 10 de registro de palabras relacionadas almacena la consulta de búsqueda recibida en la BD 12a de registro de consultas de búsqueda. En el caso en el que la condición en la etapa S2 se satisface, el servidor 10 de registro de palabras relacionadas puede registrar la consulta de búsqueda en un grupo candidato de palabras relacionadas.

60 A continuación, el servidor 10 de registro de palabras relacionadas obtiene el número de caracteres de la palabra de búsqueda (etapa S11). Específicamente, en el caso en el que la palabra de búsqueda incluida en la consulta de búsqueda es un carácter chino o similar, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas la convierte en caracteres "hiragana" o caracteres romanos. Por ejemplo, la unidad 14 de control del sistema en el servidor 10 de registro de palabras relacionadas convierte los caracteres Chinos (A 家の食卓) en fonogramas (えいけのしょくたく), aumentando de esta manera el número de caracteres de cinco caracteres a
65 nueve caracteres. Además, la unidad 14 de control del sistema en el servidor 10 de registro de palabras

relacionadas convierte los fonogramas en caracteres romanos (eikenoshokutaku), aumentando adicionalmente de esta manera el número de caracteres a 15. La unidad 14 de control del sistema en el servidor 10 de registro de palabras relacionadas obtiene el número de caracteres (por ejemplo, nueve o 15 caracteres) de la palabra de búsqueda sometida a la conversión fonográfica tal como conversión de Hiragana y conversión en caracteres romanos.

A continuación, el servidor 10 de registro de palabras relacionadas determina si el número de caracteres de la palabra de búsqueda es cinco o más (etapa S12). Específicamente, la unidad 14 de control del sistema en el servidor 10 de registro de palabras relacionadas determina si el número de caracteres de la palabra de búsqueda sometida a la conversión fonográfica es cinco o más.

En el caso en el que el número de caracteres de la palabra de búsqueda es cinco o más (Sí, en la etapa S12), el servidor 10 de registro de palabras relacionadas extrae una cadena de caracteres parcial de cuatro caracteres de la palabra de búsqueda (etapa S13). Específicamente, como se ilustra en la figura 8, la unidad 14 de control del sistema en el servidor 10 de registro de palabras relacionadas genera y extrae, de la cadena de caracteres "eikenoshokutaku" de la palabra de búsqueda, cadenas de caracteres parciales tales como una cadena de caracteres parcial p1 ("eike") y una cadena de caracteres parcial p2 ("keno") estableciendo la longitud (tamaño de ventana) L de una cadena de caracteres parcial a cuatro, y estableciendo la anchura de deslizamiento (tamaño de deslizamiento) S de una ventana a dos. Además, en base a la longitud L de una cadena de caracteres parcial, el umbral (cinco caracteres) en la etapa S2 es establece a un valor mayor que la longitud "L" de la cadena de caracteres parcial.

Como se ha descrito anteriormente, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de generación de cadenas de caracteres parciales que genera una cadena de caracteres parcial a partir de la cadena de caracteres de una palabra de búsqueda. El servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de generación de cadenas de caracteres parciales que establece la longitud de una cadena de caracteres parcial en base al número de caracteres de la palabra de búsqueda y genera una cadena de caracteres parcial que tiene la longitud establecida. El servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de generación de cadenas de caracteres parciales que establece una anchura de desplazamiento del número de caracteres de la palabra de búsqueda en el momento de la generación de una cadena de caracteres parcial en base al número de caracteres de la palabra de búsqueda y genera una cadena de caracteres parcial en la anchura de desplazamiento establecida. El servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de generación de cadenas de caracteres parciales que genera una cadena de caracteres parcial de una palabra de búsqueda obtenida mediante la conversión fonográfica de una palabra de búsqueda.

A continuación, el servidor 10 de registro de palabras relacionadas extrae una cadena de caracteres candidata de una palabra relacionada parcialmente correspondiente con una cadena de caracteres parcial de un grupo candidato de palabras relacionadas (etapa S14). Específicamente, la unidad 14 de control del sistema en el servidor 10 de registro de palabras relacionadas genera una consulta de búsqueda = "eike" OR "keno" OR "nosh" OR "shok" OR "okut" OR "utak" de la cadena de caracteres parcial ilustrada en la figura 8, y extrae cadenas de caracteres candidatas de la palabra relacionada de la BD 12b de candidatos de palabras relacionadas en base a esta consulta de búsqueda. Es decir, la unidad 14 de control del sistema en el servidor 10 de registro de palabras relacionadas extrae una cadena de caracteres candidata parcialmente correspondiente a cada una de las cadenas de caracteres parciales extraídas. Aquí, la correspondencia parcial representa, por ejemplo, en el caso de la cadena de caracteres parcial "eike", una cadena de caracteres candidata que incluye "eike" o cadenas de caracteres candidatas que incluyen "eik" como parte de "eike". De tal manera, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de extracción de cadenas de caracteres candidatas que extrae una cadena de caracteres candidata de un grupo candidato de palabras relacionadas en base a la cadena de caracteres parciales generada.

En el caso en el que el número de caracteres de la palabra de búsqueda no es igual o mayor de cinco (NO en la etapa S12), el servidor 10 de registro de palabras relacionadas extrae una cadena de caracteres candidata de una palabra relacionada parcialmente correspondiente con la palabra de búsqueda del grupo candidato de palabras relacionadas (etapa S15). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas no extrae una cadena de caracteres parcial pero extrae la propia palabra de búsqueda recibida o una cadena de caracteres candidata parcialmente correspondiente con la palabra de búsqueda convertida en fonograma de la BD 12b de candidatos de palabras relacionadas.

Además, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas puede extraer una cadena de caracteres candidata similar a una cadena de caracteres parcial o una palabra de búsqueda en base a la similitud. Además, como la similitud puede usarse la distancia entre cadenas de caracteres tales como la distancia Jaro-Winkler o la distancia Levenshtein.

A continuación, el servidor 10 de registro de palabras relacionadas calcula una puntuación de idoneidad de cadenas de caracteres candidatas de cada palabra relacionada (etapa S16). Específicamente, la unidad 14 de control del

sistema del servidor 10 de registro de palabras relacionadas calcula una puntuación de idoneidad de cada una de las cadenas de caracteres candidatas extraídas por medio de una subrutina de puntuación de idoneidad que se describirá más adelante.

5 A continuación, el servidor 10 de registro de palabras relacionadas obtiene una clasificación en el orden de las puntuaciones de idoneidad (etapa S17). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas realiza un proceso para disponer las cadenas de caracteres candidatas en orden descendente según las puntuaciones de idoneidad. La unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas extrae cadenas de caracteres candidatas del número predeterminado de puntuaciones más altas, por ejemplo, las cincuenta cadenas de caracteres candidatas de puntuación más alta y estrecha las cadenas de caracteres candidatas. De tal manera, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de clasificación que clasifica las cadenas de caracteres candidatas en orden de puntuación.

15 A continuación, el servidor 10 de registro de palabras relacionadas genera una línea de referencia para las puntuaciones de idoneidad (etapa S18). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina y genera una línea de referencia a partir de las palabras cuyas puntuaciones de idoneidad están en los primeros lugares de la clasificación en las cadenas de caracteres candidatas superiores extraídas de la palabra relacionada. Más específicamente, como se ilustra en la figura 9A, en un gráfico que ilustra puntuaciones de idoneidad con respecto a la clasificación, la unidad 14 de control del sistema establece, como una línea de referencia L1, una función aproximada que aproxima puntos de, por ejemplo, la primera a la n-ésima puntuaciones de idoneidad.

25 Aquí, la función aproximada como un ejemplo de la línea de referencia incluye una función logarítmica, una función exponencial, una función de potencia tal como una función lineal, una función cuadrática, o una función cúbica, una función de una sección cónica tal como una curva hiperbólica, y similares. Puede emplearse cualquier función aproximada siempre que aproxime o interpole puntos de un gráfico de puntuación por clasificación. En el momento de la obtención de una función aproximada, es suficiente usar un criterio del procedimiento de mínimos cuadrados, o similar. La puntuación de idoneidad relativa es, por ejemplo, el valor de la puntuación de idoneidad más alta en la clasificación y es un valor obtenido dividiendo las puntuaciones. Además, en el momento de determinar la línea de referencia L1, pueden determinarse puntuaciones de idoneidad inferiores, por ejemplo, puntuaciones de quincuagésima a décima, por el procedimiento de mínimos cuadrados. Además, dado que las puntuaciones de idoneidad se disponen en orden de clasificación, la curva de puntuación por clasificación se vuelve una gráfica débilmente decreciente.

35 Como se ha descrito anteriormente, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de generación de criterio que genera una línea de referencia de puntuaciones de idoneidad en la clasificación como un criterio para determinar si una cadena de caracteres candidata está registrada como una palabra relacionada o no en base a la puntuación de idoneidad y la clasificación de las cadenas de caracteres candidatas. El servidor 10 de registro de palabras relacionadas también funciona como un ejemplo de un medio de generación de criterio que obtiene una línea de referencia mediante una función aproximada que aproxima la relación entre la clasificación y la puntuación de idoneidad. El servidor 10 de registro de palabras relacionadas también funciona como un ejemplo del medio de generación de criterio que obtiene una línea de referencia a partir de los datos de clasificación en intervalo predeterminado.

45 A continuación, el servidor 10 de registro de palabras relacionadas determina si una divergencia entre cada una de las puntuaciones de idoneidad y la línea de referencia es igual o mayor que un umbral (etapa S19). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina si la diferencia entre la línea de referencia L1 y la puntuación más alta en la clasificación es igual a o mayor que el umbral. Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas calcula la diferencia restando un valor obtenido mediante la sustitución del valor de la clasificación en la ecuación de la línea de referencia L1 de la puntuación de idoneidad de la mayor puntuación en la clasificación. Cuando la diferencia es igual a o mayor que el umbral θ , la cadena de caracteres candidata se extrae como una cadena de caracteres de registro de la palabra relacionada. Aquí, el umbral se obtiene preliminarmente, por ejemplo, como 0,1 veces el de la línea de referencia realizando una simulación a la vez que se cambia un parámetro o similar para ajustar la puntuación de idoneidad. En este caso, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas extrae una cadena de caracteres candidata cuya puntuación de idoneidad es 1,1 veces la de la línea de referencia como una palabra relacionada.

60 Como se ha descrito anteriormente, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de extracción de cadenas de caracteres de registro que extrae una cadena de caracteres de registro para registrar, como una palabra relacionada, una cadena de caracteres candidata cuya puntuación de idoneidad está al margen de la línea de referencia L1 por el umbral predeterminado θ o mayor.

65 En el caso en el que la divergencia entre cada una de las puntuaciones de idoneidad y la línea de referencia es igual a o mayor que el umbral (SÍ, en la etapa S19), el servidor 10 de registro de palabras relacionadas registra la cadena de caracteres candidata de palabras relacionadas cuya divergencia desde la línea de referencia es igual a o mayor

que el umbral como una palabra relacionada. Específicamente, en el caso en el que la divergencia entre cada una de las puntuaciones de idoneidad y la línea de referencia es igual a o mayor que el umbral, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas registra la cadena de caracteres candidata que tiene la puntuación más alta en la clasificación en la BD 12c de palabras relacionadas como una cadena de caracteres de registro de la palabra relacionada de la palabra de búsqueda recibida. Además, en el caso en el que hay una pluralidad de (por ejemplo, tres) cadenas de caracteres candidatas cuya puntuación de idoneidad está separada de la línea de referencia por el umbral o más, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas puede almacenar, no necesariamente, la cadena de caracteres candidata como la más alta en la clasificación, las tres cadenas de caracteres candidatas más altas en la clasificación, el número predeterminado de cadenas de caracteres candidatas como las cadenas características candidatas más alta y de tercer orden, o una cadena de caracteres candidata en un orden predeterminado, tal como la cadena característica candidata de segundo orden en la clasificación. De tal manera, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo del medio de registro de palabras relacionadas que registra la cadena de caracteres de registro extraída y la palabra de búsqueda como palabras relacionadas en la base de datos de palabras relacionadas.

En el caso en el que la diferencia no es igual a o mayor que el umbral (NO en la etapa S19), el servidor 10 de registro de palabras relacionadas finaliza el proceso.

Ahora se describirá el usuario de la base de datos de palabras relacionadas construida.

En el momento de realizar una búsqueda en base a una consulta de búsqueda recibida desde el terminal 30 del usuario, el servidor 20 de suministro de información como el medio de extracción de palabras relacionadas extrae una palabra relacionada correspondiente a una palabra de búsqueda en la consulta de búsqueda con referencia a la BD 12c de palabras relacionadas en el servidor 10 de registro de palabras relacionadas. Por ejemplo, el servidor 20 de suministro de información transmite la consulta de búsqueda recibida al servidor 10 de registro de palabras relacionadas. El servidor 10 de registro de palabras relacionadas recibe la consulta de búsqueda del servidor 20 de suministro de información y extrae una cadena de caracteres de registro correspondiente a la palabra de búsqueda en la consulta de búsqueda de la BD 12c de palabras relacionadas. El servidor 10 de registro de palabras relacionadas como un ejemplo del medio de salida de palabras relacionadas transmite, como una palabra relacionada, la cadena de caracteres de registro extraída al servidor 20 de suministro de información. El servidor 20 de suministro de información transmite la palabra relacionada recibida al terminal 30. Como se ilustra en la figura 6, el terminal 30 muestra la palabra relacionada recibida como "¿tal vez "XXX"?" en un cuadro 43 de presentación de palabra relacionada. Además, una pluralidad de palabras relacionadas puede aparecer como "palabra relacionada en la clasificación 1" y "palabra relacionada en la clasificación 2". No sólo en el caso en el que el número de resultados de búsqueda es cero o pequeño, puede visualizarse una palabra relacionada.

(3.3 Subrutina de cálculo de puntuación de idoneidad)

Ahora se describirá una subrutina de cálculo de una puntuación de idoneidad usando las figuras 10 y 11.

Después de extraer la cadena característica candidata de la palabra relacionada en la etapa S14 o S15, como se ilustra en la figura 10, el servidor 10 de registro de palabras relacionadas calcula la distancia entre la cadena de caracteres candidata de la palabra relacionada extraída y la palabra de búsqueda (etapa S30). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas calcula, por ejemplo, una distancia (w, u) entre una palabra de búsqueda "u" y una cadena de caracteres candidata "w" mediante la ecuación (1).

[Ecuación 1]

$$\text{Distancia}(w, u) = \alpha \text{Dist}(w, u) + \beta \text{Dist. Yomi}(w, u) \dots (1)$$

donde "la distancia (w, u)" representa la distancia entre la palabra de búsqueda "u" y la cadena de caracteres candidata "w" y es un índice (un ejemplo de similitud que indica cómo de similares son la palabra de búsqueda "u" y la cadena de caracteres candidata "w". Dist(w,u) representa la distancia Jaro-Winkler entre la cadena de caracteres candidata "w" y la palabra de búsqueda "u", Dist.Yomi(w,u) representa la distancia Jaro-Winkler entre la lectura de la cadena de caracteres candidata "w" y la palabra de búsqueda "u". Los coeficientes α y β tienen la relación de $\alpha + \beta = 1$. Cuando se desea dar importancia a la lectura, la relación del coeficiente β se aumenta (por ejemplo, $\alpha < \beta$, $\alpha = 0,3$, $\beta = 0,7$). Aunque "distancia(w,u)" es preferiblemente la distancia Jaro-Winkler que puede expresarse de 0 a 1 como una distancia, no se limita a la distancia Jaro-Winkler sino que puede ser la distancia Levenshtein o un índice que expresa la similitud entre la palabra de búsqueda "u" y la cadena de caracteres candidata "w".

A continuación, el servidor 10 de registro de palabras relacionadas obtiene la frecuencia de aparición de una cadena de caracteres candidata de cada palabra relacionada (etapa S31). En concreto, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas calcula la frecuencia de aparición de una cadena de caracteres candidata de cada palabra relacionada en un registro de consulta de búsqueda. Más específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas calcula P(w) para cada cadena de

caracteres candidata "w" en base a los datos de la BD 12a de registro de consultas de búsqueda mediante la ecuación (2).

[Ecuación 2]

$$P(w) = \log_{10}(\text{Occ. Count}(w)) \dots\dots(2)$$

donde P (w) se refiere a la frecuencia de ocurrencia que indica la frecuencia con que la cadena de caracteres candidata "w" es buscada en los datos de registro de búsqueda en la BD 12a de registro de consultas de búsqueda. Además, Occ.Count(w) representa el número de apariciones de la cadena de caracteres candidata "w" (el número de veces de uso de la cadena de caracteres candidata) en el registro de consultas de búsqueda. En el caso de usar los datos del registro de consultas de búsqueda en la BD 12a de registro de consultas de búsqueda, una palabra que tiene un alto P(w) es una palabra común que se introduce a menudo en las búsquedas por diversos usuarios. De tal manera, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas como un ejemplo de un medio de cálculo de puntuación de idoneidad, calcula una puntuación de idoneidad en base a los datos en un medio de almacenamiento de consultas de (BD 12a de registro de consultas de búsqueda).

A continuación, el servidor 10 de registro de palabras relacionadas calcula la buscabilidad (disponibilidad) a partir del número de búsquedas en una BD 22a de productos de una cadena de caracteres candidata de cada palabra relacionada (etapa S32). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas transmite una cadena de caracteres candidata "w" de cada palabra relacionada al servidor 20 de suministro de información, y la unidad de control del sistema 24 del servidor 20 de suministro de información obtiene el número de búsquedas (el número de resultados de búsqueda) en la BD 22a de productos y transmite el número de búsquedas de cada una de las cadenas de caracteres candidatas "w" al servidor 10 de registro de palabras relacionadas.

La unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas calcula el valor de la buscabilidad de la cadena de caracteres candidata "w" a partir del número de búsquedas mediante la ecuación (3).

[Ecuación 3]

$$\text{Disponibilidad}(w) = \begin{cases} 1 & (\text{n}^\circ \text{ de ResultadoBúsqueda}(w) > 0) \\ 0 & (\text{n}^\circ \text{ de ResultadoBúsqueda}(w) = 0) \end{cases} \dots\dots(3)$$

"Disponibilidad(w)" representa un coeficiente de peso para el número de resultados de búsqueda (el número de aciertos de búsqueda) de búsquedas con la cadena de caracteres candidata "w" y se expresa por la siguiente ecuación. "nº ResultadoBúsqueda (w)" representa el número de resultados de búsqueda de las cadenas de caracteres candidatas "w". La disponibilidad incluye no sólo la ecuación (3) como se ilustra en la figura 11A, sino también variaciones como se ilustra en las figuras 11B y 11C.

A continuación, el servidor 10 de registro de palabras relacionadas calcula una puntuación de idoneidad de la cadena de caracteres candidata de cada palabra relacionada a partir de la distancia entre la palabra de búsqueda y la cadena de caracteres candidata de la palabra relacionada, la frecuencia de aparición de la cadena de caracteres candidata de cada palabra relacionada, y la buscabilidad de la cadena de caracteres candidata de cada palabra relacionada (etapa S33). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas calcula la puntuación de idoneidad de cada cadena de caracteres candidata "w" mediante la ecuación (4).

[Ecuación 4]

$$\text{Puntuación}(w|u) = \frac{P(w) + \gamma}{\text{Distancia}(w,u) + \delta} \cdot \text{Disponibilidad}(w) \dots\dots(4)$$

donde γ y δ son parámetros para ajustar la puntuación de idoneidad. El parámetro γ es un valor de ajuste para P(w) y expresa el grado de autorización de corrección para una palabra que no se introduce demasiado. El valor P(w) de una palabra que no se introduce demasiado es pequeño, de manera que es difícil captar una nueva palabra y un nuevo producto por una búsqueda. En el caso en el que se desea captar una palabra que no se usa demasiado por una búsqueda, el valor del parámetro γ se aumenta mediante un ajuste en el lado del servidor.

El parámetro δ es un valor de ajuste para Distancia(w,u) y principalmente impide que Distancia(w,u) se vuelva cero e impide que el valor de puntuación diverja. Además, el servidor 10 de registro de palabras relacionadas preajusta los parámetros γ y δ en la ecuación (4) y el parámetro α y β en la ecuación (1) (por ejemplo, $\gamma > \delta$, $\gamma > 1$, $\delta < 1$, $\gamma = 2$, $\delta = 0,01$).

Como se ha descrito anteriormente, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de

5 un medio de cálculo de puntuación que calcula la puntuación de idoneidad de una cadena de caracteres candidata en base a la similitud entre la cadena de caracteres candidata y la palabra de búsqueda, el número de veces de uso de la cadena de caracteres candidata, y el número de resultados de búsqueda de una búsqueda por la cadena de caracteres candidata. El servidor 10 de registro de palabras relacionadas funciona como un medio de ajuste de parámetros que ajusta la información de similitud y el número de veces de uso.

10 En la implementación a modo de ejemplo anterior, un grupo candidato de palabras relacionadas obtenido extrayendo cadenas de caracteres candidatas de una palabra relacionada cuyo número de resultados de búsqueda es uno o más (un ejemplo de una condición predeterminada) de la BD 12a de registro de consultas de búsqueda en la que se almacenan los registros relativos a una consulta de búsqueda, es generado y almacenado en la BD 12b de candidatos de palabras relacionadas. Se recibe una consulta de búsqueda de una palabra de búsqueda introducida por el usuario, y se genera una cadena de caracteres parcial a partir de una cadena de caracteres de la palabra de búsqueda. En base a las cadenas de caracteres parciales, se extrae una cadena de caracteres candidata del grupo candidato de palabras relacionadas en la BD 12b de candidatos de palabras relacionadas. Se calcula una puntuación de idoneidad de la cadena de caracteres candidata en base a la similitud entre la cadena de caracteres candidata y la palabra de búsqueda, el número de veces de uso de la cadena de caracteres candidata, y el número de resultados de búsqueda de la búsqueda de la cadena de caracteres candidata. Las cadenas de caracteres candidatas se clasifican por orden de puntuación. Se genera una línea de referencia de una puntuación de idoneidad para la clasificación como un criterio para determinar si la cadena de caracteres candidata está registrada como una palabra relacionada o no en base a la puntuación de idoneidad y la clasificación de la cadena de caracteres candidata. Como una cadena de caracteres de registro a registrar como una palabra relacionada, se extrae una cadena de caracteres candidata cuya puntuación de idoneidad está separada de la línea de referencia por un umbral predeterminado o mayor. La cadena de caracteres de registro extraída se registra como una palabra relacionada de la palabra de búsqueda en la BD 12c de palabras relacionadas. En consecuencia, como se extrae una cadena de caracteres candidata del grupo candidato de palabras relacionadas en base a una cadena de caracteres parcial que tiene un menor número de caracteres, puede suprimirse el registro de una palabra relacionada innecesaria mediante la puntuación de idoneidad de la cadena de caracteres candidata, y puede mejorarse la precisión de una palabra relacionada. Además, de acuerdo con la realización, la velocidad del proceso puede aumentarse.

30 Cuando una palabra de búsqueda se presenta al usuario en base a una palabra relacionada, el usuario puede conseguir un objeto de búsqueda objetivo rápidamente.

35 En el caso en el que el servidor 10 de registro de palabras relacionadas establece la longitud de una cadena de caracteres parcial en base al número de caracteres de una palabra de búsqueda y genera una cadena de caracteres parcial que tiene la longitud establecida, determinando la longitud de la cadena de caracteres parcial, la eficiencia del proceso puede mejorarse, y la velocidad de procesamiento y la precisión de una palabra relacionada pueden mejorarse.

40 En el caso en el que el servidor 10 de registro de palabras relacionadas establece una anchura de desplazamiento del número de caracteres de una palabra de búsqueda en el momento de la generación de una cadena de caracteres parcial en base al número de caracteres de una palabra de búsqueda y genera una cadena de caracteres parcial con la anchura de desplazamiento establecida, la velocidad del proceso puede aumentarse.

45 En el caso en el que el servidor 10 de registro de palabras relacionadas genera una cadena de caracteres parcial para una palabra de búsqueda obtenida por conversión fonográfica de una palabra de búsqueda, incluso en el caso de una palabra corta, puede extraerse eficientemente una cadena de caracteres parcial, y la velocidad de procesamiento y la precisión de una palabra relacionada pueden mejorarse.

50 En el caso en el que el servidor 10 de registro de palabras relacionadas incluye adicionalmente, como un medio de ajuste de parámetro, los parámetros δ y γ que ajustan la información de la similitud y el número de veces de uso del medio de cálculo de puntuación, ajustando los parámetros δ y γ , la puntuación de idoneidad puede cambiarse de acuerdo con las características y situaciones del usuario, y la precisión de una palabra relacionada puede mejorarse.

55 El servidor 10 de registro de palabras relacionadas como medio de generación de referencia de determinación obtiene la línea de referencia mediante una función aproximada que aproxima la relación entre la clasificación y la puntuación de idoneidad, de manera que una cadena de caracteres de registro de una palabra relacionada se detecta más fácilmente como un valor anormal fuera de la función aproximada, y la precisión de la palabra relacionada puede mejorarse más.

60 En el caso en el que el servidor 10 de registro de palabras relacionadas como medio de generación de referencia de determinación obtiene la línea de referencia L1 por los datos de la clasificación en un intervalo predeterminado (por ejemplo, clasificaciones superiores), puede detectarse una cadena de caracteres candidata desviada de la línea de referencia L1, de manera que puede extraerse apropiadamente una cadena de caracteres candidata en un rango superior.

65 En el caso en el que se proporciona adicionalmente la BD 22a de productos como un ejemplo de un medio de

almacenamiento de palabras de búsqueda que almacena una palabra de búsqueda relacionada con un producto que se vende en un sitio EC interior o exterior, particularmente, la disponibilidad puede calcularse fácilmente usando datos de la BD 22a de productos y puede reflejarse en una puntuación de idoneidad.

5 En el caso en el que el servidor 10 de registro de palabras relacionadas incluye adicionalmente la BD 12a de registro de consultas de búsqueda como un ejemplo de un medio de almacenamiento de consulta de búsqueda que almacena un registro en una consulta de búsqueda y calcula la puntuación de idoneidad en base a los datos de la BD 12a de registro de consultas de búsqueda, cuando las cadenas de caracteres candidatas se extraen de palabras almacenadas en la BD 12a de registro de consultas de búsqueda, las cadenas de caracteres candidatas pueden estrecharse monodimensionalmente para generar cadenas de caracteres candidatas en las que se concentran los intereses del usuario. En particular, en el caso de usar un registro reciente, puede extraerse una cadena de caracteres candidata en la que las preferencias del usuario se reflejan particularmente.

15 El servidor 10 de registro de palabras relacionadas puede establecer el límite superior al número de cadenas de caracteres de registro extraídas por el medio de extracción de cadenas de caracteres de registro en la etapa S19. En este caso, el registro de palabras relacionadas innecesarias puede suprimirse, y la precisión de las palabras relacionadas puede mejorarse.

20 En el caso en el que el servidor 20 de suministro de información extrae y transmite una palabra relacionada correspondiente a una palabra de búsqueda de una consulta de búsqueda recibida con referencia a la base de datos de palabras relacionadas, puede presentarse rápidamente al usuario una palabra relacionada de alta precisión mediante la cual el usuario puede alcanzar un objeto de búsqueda objetivo.

25 Además, la línea de referencia puede determinarse también a partir de una cadena de caracteres candidata cuya puntuación de idoneidad está en una clasificación inferior que una cadena de caracteres candidata de una palabra relacionada de alto orden extraída. Más específicamente, como se ilustra en la figura 9B, en un gráfico que ilustra puntuaciones de idoneidad en relación con la clasificación, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina, por ejemplo, una línea recta (línea de referencia L2) que conecta las puntuaciones de idoneidad en el 50° orden y el 10° orden.

30 La línea de referencia L2 se expresa mediante una ecuación de una línea recta de $y = ax + b$ en el gráfico en el que el eje x indica la clasificación y el eje y indica la puntuación de idoneidad relativa. Aquí, una puntuación de idoneidad relativa es, por ejemplo, un valor de una puntuación de idoneidad en el primer orden en la clasificación, y es un valor obtenido dividiendo cada puntuación. Además, en el momento de determinar la línea de referencia L2, pueden determinarse puntuaciones de idoneidad de bajo orden, es decir, pueden determinarse mediante el procedimiento de mínimos cuadrados puntuaciones de 50° orden al 10° orden. Aquí, como se ilustra en la figura 9B, las puntuaciones de idoneidad inferior se aproximan entre sí en el orden de clasificación en comparación con las puntuaciones de idoneidad superiores.

40 Aquí, la línea de referencia L2 es un ejemplo de la función aproximada que aproxima la relación entre la clasificación y la puntuación de idoneidad y una línea recta que conecta las puntuaciones de idoneidad en el 50° orden y el 10° orden. Además, la función aproximada no se limita a la línea recta (función lineal), sino que puede ser una función logarítmica, una función exponencial, una función de potencia tal como una función cuadrática o una función cúbica, una función de una sección cónica, tal como una curva hiperbólica, y similares.

45 Después de determinar la línea de referencia L2, el servidor 10 de registro de palabras relacionadas obtiene el valor de una intersección "b" al eje "y" de la línea de referencia (el eje de la puntuación de idoneidad relativa) y establece un valor obtenido multiplicando el valor de la intersección "b" con una constante como umbral θ .

50 Además, en la etapa S13, el servidor 10 de registro de palabras relacionadas puede extraer una cadena de caracteres parcial y, de acuerdo con el número de caracteres de una palabra de búsqueda, cambiar dinámicamente la longitud L de la cadena de caracteres parcial y la anchura de deslizamiento S de la ventana. Por ejemplo, en el caso en el que el número de caracteres de una palabra de búsqueda es grande, la longitud L de la cadena de caracteres parcial se aumenta, y la anchura de deslizamiento S se ensancha. En el caso en el que L se ajusta a 4 y S se ajusta a 3 como valores de referencia para el número de caracteres de 15, cuando el número de caracteres de una palabra de búsqueda es 40, la longitud de la cadena de caracteres se aumenta como $L = 8$ y $S = 3$, la anchura de deslizamiento se aumenta como $L = 4$ y $S = 5$, y la longitud de la cadena de caracteres y la anchura de deslizamiento se aumentan como $L = 8$ y $S = 5$. En el caso en el que el servidor 10 de registro de palabras relacionadas cambia dinámicamente la longitud L de una cadena de caracteres parcial y la anchura de deslizamiento S de ventana de acuerdo con el número de caracteres de una palabra de búsqueda, la velocidad de procesamiento y la precisión de una palabra relacionada pueden mejorarse.

65 Un sitio de compras puede ser un sitio de EC exterior conectado a la red 3, y el servidor 20 de suministro de información no se limita a la búsqueda de productos, sino que puede ser un servidor de búsqueda general. En este caso, la facilidad de búsqueda (disponibilidad) se calcula a partir del número de veces de búsqueda de un servidor de búsqueda en un sitio EC exterior o similar.

[4. Funcionamiento de un modo de realización en un sistema de registro de palabras relacionadas]

- 5 El funcionamiento de un modo de realización en el sistema 1 de registro de palabras relacionadas se describirá usando la figura 6 y las figuras 12 a 16 de acuerdo con un modo de realización de la presente invención. Se usan los mismos números de referencia para las operaciones que son las mismas o similares a las de la implementación a modo de ejemplo, y se describirán principalmente los puntos diferentes. Los otros modos de realización y modificaciones también se describirán de una manera similar.
- 10 (4.1 Generación de registro de consultas de búsqueda y grupo de conjuntos de cadenas de caracteres)
- En primer lugar, se describirá usando la figura 6 y las figuras 12 a 14 la generación de un registro de consultas de búsqueda y un grupo de conjuntos de cadenas de caracteres.
- 15 Además, la descripción se iniciará desde un estado donde la BD 12a de registro de consultas de búsqueda está construida en cierta medida. En la BD 12a de registro de consultas de búsqueda, se almacena una consulta de búsqueda asociada con la hora de recepción de una consulta de búsqueda junto con información de identificación de usuario. Además, la consulta de búsqueda se almacena de manera que está asociada también con la información de categorías de un producto (la información de categorías se describirá más adelante).
- 20 Por ejemplo, en el caso de una consulta de búsqueda desde el terminal 30 del usuario A que inicia sesión en un sitio de compras, ejemplos de la información de identificación del usuario incluyen el ID de usuario del usuario A y la dirección IP existente en el encabezado de la consulta de búsqueda transmitida desde el terminal 30 del usuario A.
- 25 Ejemplos de la hora de recepción incluyen la hora de recepción cuando el servidor 20 de suministro de información recibe la consulta de búsqueda del terminal 30 y la hora de recepción cuando el servidor 10 de registro de palabras relacionadas recibe la consulta de búsqueda del servidor 20 de suministro de información. En el caso de la hora de recepción cuando el servidor 20 de suministro de información recibe la consulta de búsqueda del terminal, el servidor 20 de suministro de información transmite también la hora de recepción de la consulta de búsqueda al servidor 10 de registro de palabras relacionadas. La hora de recepción puede ser un sello temporal fijado cuando se almacena la consulta de búsqueda en la BD 12a de registro de consultas de búsqueda.
- 30 En primer lugar, el usuario (por ejemplo, el usuario A) inicia sesión en un sitio de compras mediante el terminal 30 del usuario A, como se ilustra en la figura 6, se transmite información de la página web 40 como la página principal del sitio de compras desde el servidor 20 de suministro de información al terminal 30. La página web 40 se muestra en la unidad 33 de visualización del terminal 30. En la página web 40 como la página principal del sitio de compras, se muestra una vista 45 de categorías principales y se vincula a las categorías. Además, se muestra una vista 46 de categorías de los productos y servicios de compañías de grupos relacionados del sitio de compras y se vincula a las categorías. En la página web 40, se muestra información del usuario A, "¡Hola, A!".
- 40 Cuando se hace clic en el enlace de la categoría de producto de "electrónica del hogar" en la página web 40 como se ilustra en la figura 6, se muestra una página web 50 de "electrónica del hogar" en la unidad 33 de visualización como se ilustra en la figura 12. En la página web 50, se muestra una lista de subcategorías 56 en una lista de categorías 55 de la categoría principal "electrónica del hogar".
- 45 Cuando se introduce una palabra de búsqueda "xxx" en un cuadro 51 de entrada de palabra de búsqueda en la página web 50 y se hace clic en un botón de búsqueda 52, la unidad 36 de control del sistema del terminal 30 transmite la consulta de búsqueda al servidor 20 de suministro de información a través de la unidad 31 de comunicación. La consulta de búsqueda tiene la palabra de búsqueda introducida, la información de categorías (número de identificación de categoría o similar, indicativo de una lista de categorías 65 de la categoría principal "electrónica del hogar") de la página web 50, el ID de usuario del (usuario A) como un ejemplo de la información de identificación de usuario, y similar.
- 50 A continuación, el servidor 20 de suministro de información recibe la consulta de búsqueda del terminal 30, la transmite al servidor 10 de registro de palabras relacionadas, y busca un producto en la BD 22a de productos. El servidor 20 de suministro de información especifica la dirección IP del emisor de la consulta de búsqueda y transmite la dirección IP también al servidor 10 de registro de palabras relacionadas. La dirección IP es un ejemplo de la información de identificación del usuario y se usa en lugar del ID de usuario cuando el usuario ha iniciado sesión.
- 55 A continuación, como se ilustra en la figura 13, el servidor 10 de registro de palabras relacionadas recibe la consulta de búsqueda que incluye la palabra de búsqueda (etapa S41). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas recibe la consulta de búsqueda del servidor 20 de suministro de información a través de la unidad 11 de comunicación, y almacena la consulta de búsqueda recibida junto con la hora de recepción de la consulta de búsqueda en la BD 12a de registro de consultas de búsqueda. Además, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas almacena la consulta de búsqueda junto con la dirección IP en la BD 12a de registro de consultas de búsqueda. Además, la unidad 14 de
- 60
- 65

control del sistema del servidor 10 de registro de palabras relacionadas puede poner la dirección IP en la consulta de búsqueda. Como es suficiente con conocer el orden de recepción de las consultas de búsqueda, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas puede almacenar consultas de búsqueda asociadas a órdenes de recepción, no la hora de recepción, en la BD 12a de registro de consultas de búsqueda. De tal manera, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de almacenamiento de consultas de búsqueda que almacena consultas de búsqueda recibidas en orden de recepción.

A continuación, el servidor 10 de registro de palabras relacionadas determina si una consulta de búsqueda anterior con un ID de usuario correspondiente existe o no en el registro de consultas de búsqueda (etapa S42). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas extrae una consulta de búsqueda cuyo ID de usuario corresponde con el ID de usuario del usuario A y cuya hora de recepción es anterior a la de la consulta de búsqueda recibida (un ejemplo de orden de recepción anterior) desde la BD 12a de registro de consultas de búsqueda. En el caso en el que la consulta de búsqueda anterior puede extraerse, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina que la consulta de búsqueda anterior existe en el registro de consultas de búsqueda. En el caso en el que la consulta de búsqueda anterior no puede extraerse, la unidad 14 de control del sistema determina que no existe una consulta de búsqueda anterior. Además, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas puede extraer una consulta de búsqueda anterior cuya dirección IP concuerda con la dirección IP de la consulta de búsqueda recibida y cuya hora de recepción es anterior a la de la consulta de búsqueda recibida desde la BD 12a de registro de consultas de búsqueda.

Aquí, el caso en el que los IDs de usuario concuerdan o el caso en el que las direcciones IP concuerdan es un ejemplo de una condición de extracción de consultas de búsqueda predeterminada. De tal manera, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de extracción de consultas de búsqueda que extrae una consulta de búsqueda anterior cuya información de identificación de usuario concuerda con la del usuario que introdujo la consulta de búsqueda como la condición de extracción de consultas de búsqueda.

En el caso en el que una consulta de búsqueda anterior que tiene el ID de usuario coincidente existe en el registro de consultas de búsqueda (SÍ, en la etapa S42), el servidor 10 de registro de palabras relacionadas determina si la hora de recepción y la hora de recepción de la consulta de búsqueda anterior están o no dentro de un tiempo predeterminado (etapa S43). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas extrae una consulta de búsqueda anterior cuya hora de recepción tiene una diferencia con la de la consulta de búsqueda recibida, que está dentro de un tiempo predeterminado (por ejemplo, tres segundos a 60 segundos). En el caso en el que la consulta de búsqueda anterior puede extraerse, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina que la consulta de búsqueda anterior se recibió dentro del tiempo predeterminado. En el caso en el que la consulta de búsqueda anterior no puede extraerse, la unidad 14 de control del sistema determina que no se recibió ninguna consulta de búsqueda anterior dentro del tiempo predeterminado.

Aquí, en este caso, el tiempo predeterminado es un ejemplo de la condición de extracción de consultas de búsqueda predeterminada. De tal manera, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de extracción de consultas de búsqueda que extrae una consulta de búsqueda anterior cuya hora de recepción es anterior a la de una consulta de búsqueda recibida en base a la condición de extracción de consultas de búsqueda predeterminada. El servidor 10 de registro de palabras relacionadas también funciona como un medio de extracción de consultas de búsqueda que extrae una consulta de búsqueda anterior cuya hora de recepción está dentro del tiempo predeterminado desde la hora de recepción de la consulta de búsqueda como la condición de extracción de consultas de búsqueda.

En el caso en el que la hora de recepción y la hora de recepción de la consulta de búsqueda anterior están dentro del tiempo predeterminado (SÍ en la etapa S43), el servidor 10 de registro de palabras relacionadas almacena un conjunto de cadenas de caracteres de la palabra de búsqueda recibida y la palabra de búsqueda anterior cuyo tiempo de recepción es más cercano (etapa S44). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas selecciona una palabra de búsqueda que construye la consulta de búsqueda recibida y una palabra de búsqueda que construye la consulta de búsqueda anterior cuya hora de recepción es más cercana a las consultas anteriores extraídas y almacena un conjunto de cadenas de caracteres de la palabra de búsqueda anterior y la palabra de búsqueda recibida en la BD 12b de conjuntos de cadenas de caracteres.

Además, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas puede estrechar las consultas de búsqueda anteriores cuya información de categoría (por ejemplo, la categoría principal) concuerda como un ejemplo de la condición de extracción de consultas de búsqueda predeterminada.

A continuación, el servidor 10 de registro de palabras relacionadas almacena la consulta de búsqueda junto con el ID de usuario y la hora de recepción en el registro de consultas de búsqueda (etapa S45). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas almacena la consulta de búsqueda asociada con la información de categorías, el ID de usuario (o la dirección IP), y la hora de recepción en la BD 12a

de registro de consultas de búsqueda. En el caso en el que no exista una consulta de búsqueda anterior cuyo ID de usuario concuerde en el registro de consultas de búsqueda (NO en la etapa S42) o en el caso en el que la hora de recepción de la consulta de búsqueda y la hora de recepción de la consulta de búsqueda anterior no estén dentro del tiempo predeterminado (NO en la etapa S43), la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas almacena la consulta de búsqueda asociada al identificador de usuario y la hora de recepción en la BD 12a de registro de consultas de búsqueda.

De tal manera, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de almacenamiento de conjuntos de cadenas de caracteres que almacena, como un conjunto de cadenas de caracteres, una palabra de búsqueda anterior que construye una consulta de búsqueda anterior extraída y una palabra de búsqueda que construye una consulta de búsqueda recibida.

A continuación, el servidor 20 de suministro de información realiza una búsqueda de productos y, posteriormente, transmite, como un resultado de búsqueda, la información de una página web 60 al terminal 30 como se ilustra en la figura 14.

A continuación, el terminal 30 recibe la información de la página web 60 del servidor 20 de suministro de información y la muestra en la unidad 33 de visualización.

Cuando el usuario A ve el resultado de la búsqueda de la página web 60 y encuentra un producto objetivo, el usuario A hace clic en el producto en un cuadro 63 de visualización de productos. Cuando no hay ningún producto objetivo, el usuario A introduce otra palabra de búsqueda en un cuadro 61 de entrada de palabra de búsqueda y hace clic en un botón 62 de búsqueda. Como se ha descrito anteriormente, el terminal 30 transmite una consulta de búsqueda que incluye la palabra de búsqueda "xxx" e información de la categoría ("electrónica del hogar") al servidor 20 de suministro de información. El servidor 20 de suministro de información recibe la consulta de búsqueda, y la transmite al servidor 10 de registro de palabras relacionadas para realizar una búsqueda. El servidor 10 de registro de palabras relacionadas recibe la consulta de búsqueda (etapa S41) y realiza los procesos en las etapas S42 a S45 como se ha descrito anteriormente.

Los procesos se realizan también en una consulta de búsqueda de otro usuario B y similares, un registro de consulta de búsqueda de consultas de búsqueda asociadas a la hora de recepción junto con información de la categoría, cada ID de usuario, o cada dirección IP se genera en la BD 12a de registro de consultas de búsqueda, y un grupo de conjuntos de cadenas de caracteres hecho de cadenas de caracteres se genera en la BD 12b de conjuntos de cadenas de caracteres.

(4.2 Registro de palabra relacionada)

A continuación, se describirá usando las figuras 15 y 16 el registro de una palabra relacionada.

En primer lugar, como se ilustra en la figura 15, el servidor 10 de registro de palabras relacionadas determina si se satisface o no una condición de inicio de extracción (etapa S50). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina si, como un ejemplo de la condición de inicio de extracción, el número de consultas de búsqueda (el número total de consultas de búsqueda) en el registro de consultas de búsqueda en la BD 12a de registro de consultas de búsqueda o el número de elementos del grupo de conjuntos de cadenas de caracteres (el número total de grupos de conjuntos de cadenas de caracteres) almacenados en la BD 12b de conjuntos de cadenas de caracteres excede un umbral predeterminado o no. La unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas también puede determinar si ha pasado o no el tiempo predeterminado desde que se satisfizo por última vez la condición de inicio de extracción de conjuntos de cadenas de caracteres.

En el caso en el que la condición de inicio de extracción se satisface (Sí, en la etapa S50), el servidor 10 de registro de palabras relacionadas extrae un conjunto de cadenas de caracteres cuya palabra de búsqueda es la misma que la palabra de búsqueda anterior del grupo de conjuntos de cadenas de caracteres (etapa S51). Específicamente, como se ilustra en la figura 16, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas extrae un conjunto de cadenas de caracteres 70 cuya palabra de búsqueda concuerda con una palabra 71 de búsqueda anterior (las palabras de búsqueda son las mismas) de la BD 12b de conjuntos de cadenas de caracteres. En el caso en el que la condición de inicio de extracción no se satisface (NO en la etapa S50), finaliza el proceso de registro de palabras relacionadas.

Además, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas puede extraer un conjunto de cadenas de caracteres cuya palabra de búsqueda es similar a la palabra de búsqueda anterior de la BD 12b de conjuntos de cadenas de caracteres de acuerdo con la condición de inicio de extracción de conjuntos de cadenas de caracteres predeterminada. Cuando las palabras de búsqueda son similares entre sí, esto significa que, por ejemplo, el valor de similitud, en base a la distancia entre las cadenas de caracteres tal como la distancia Jaro-Winkler o la distancia Levenshtein, es igual a o mayor que un umbral predeterminado. Por ejemplo, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas también extrae un conjunto de cadenas de

caracteres que tiene una palabra de búsqueda de procesamiento "xxx" similar a la palabra 71 de búsqueda anterior.

5 Como se ha descrito anteriormente, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de extracción de cadenas de caracteres que extrae un conjunto de cadenas de caracteres cuya palabra de búsqueda es igual o similar a la palabra de búsqueda anterior del medio de almacenamiento de conjuntos de cadenas de caracteres de acuerdo con la condición de inicio de extracción de conjuntos de cadenas de caracteres predeterminada.

10 A continuación, el servidor 10 de registro de palabras relacionadas determina si el número total de conjuntos de cadenas de caracteres extraídos es igual a o mayor que un número predeterminado (etapa S52). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina si el número total de conjuntos de cadenas de caracteres con respecto a la palabra 71 de búsqueda anterior es igual o mayor que un número predeterminado (por ejemplo, tres). Como se ilustra en la figura 16, como conjuntos de cadenas de caracteres con respecto a la palabra 71 de búsqueda anterior, hay un total de tres conjuntos; un conjunto 70 de cadenas de caracteres (que tiene la palabra 71 de búsqueda anterior - palabra de búsqueda 71), un conjunto 70 de cadenas de caracteres (que tiene la palabra 71 de búsqueda anterior - una palabra de búsqueda 72), y un conjunto 70 de cadenas de caracteres (que tiene la palabra 71 de búsqueda anterior - una palabra de búsqueda 73).

20 En el caso en el que el número total de conjuntos de cadenas de caracteres extraídas es igual a o mayor que un número predeterminado (SÍ, en la etapa S52), el servidor 10 de registro de palabras relacionadas determina si hay o no un conjunto de cadenas de caracteres que satisfaga la condición de registro en los conjuntos de cadenas de caracteres extraídas (etapa S53). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina si el número de los conjuntos 70 de cadenas de caracteres es igual a o mayor que un número predeterminado (por ejemplo, 10 o más) como un ejemplo de la condición de registro. Como se ha descrito anteriormente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina si el número de conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda 72, 73 y 74 en los conjuntos 70 de cadenas de caracteres cuyas palabras de búsqueda son las mismas que o similares a la palabra 71 de búsqueda anterior excede o no el umbral predeterminado. La unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas especifica, como palabras relacionadas, un conjunto de cadenas de caracteres que excede el umbral predeterminado. Además, cuando existe una palabra de búsqueda "xxx" similar a la palabra de búsqueda 73 "xxy", la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas puede contarla como el mismo conjunto de cadenas de caracteres.

35 Como se ha descrito anteriormente, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de especificación de palabras relacionadas que especifica un conjunto de cadenas de caracteres como palabras relacionadas de los conjuntos de cadenas de caracteres extraídos en base a una condición de registro predeterminada. El servidor 10 de registro de palabras relacionadas también funciona como un ejemplo de un medio de especificación de palabras relacionadas, cuando el número de conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda en un conjunto de cadenas de caracteres cuya palabra de búsqueda es la misma que o similar a una palabra de búsqueda anterior excede un umbral predeterminado como una condición de registro, que especifica el conjunto de cadenas de caracteres como palabras relacionadas.

45 La unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas también puede determinar si la relación de los conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda en un conjunto de cadenas de caracteres cuya palabra de búsqueda es la misma o similar a la palabra 71 de búsqueda anterior excede o no un umbral predeterminado (por ejemplo, 80%) como un ejemplo de una condición de registro.

50 Como se ha descrito anteriormente, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de especificación de palabras relacionadas, cuando la relación de conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda en un conjunto de cadenas de caracteres cuya palabra de búsqueda es la misma que o similar a una palabra de búsqueda anterior excede un umbral predeterminado como una condición de registro, que especifica el conjunto de cadenas de caracteres como palabras relacionadas.

55 A continuación, en el caso en el que es el conjunto de cadenas de caracteres el que satisface la condición de registro (SÍ, en la etapa S53), el servidor 10 de registro de palabras relacionadas registra el conjunto de cadenas de caracteres como palabras relacionadas (etapa S54). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas registra, como se ilustra en la figura 16, el conjunto 70 de cadenas de caracteres que tiene <la palabra 71 de búsqueda anterior - la palabra de búsqueda 72> como palabras relacionadas en la BD 12c de palabras relacionadas ya que el número de conjuntos 70 de cadenas de caracteres que tiene <la palabra 71 de búsqueda anterior - la palabra de búsqueda 71> es tres (15%), el número 70 de los conjuntos de cadenas de caracteres que tiene <la palabra 71 de búsqueda anterior - la palabra de búsqueda 72> es 16 (80%), y el número 70 del conjunto de cadenas de caracteres que tiene <la palabra 71 de búsqueda anterior - la palabra de búsqueda 73> es uno (5%).

65 Como se ha descrito anteriormente, el servidor 10 de registro de palabras relacionadas funciona como un ejemplo de un medio de registro de palabras relacionadas que registra un conjunto de cadenas de caracteres especificado como

palabras relacionadas.

Además, en el caso en el que no hay ningún conjunto de cadenas de caracteres que satisfaga la condición de registro (NO en la etapa S53), el servidor 10 de registro de palabras relacionadas determina si un conjunto de cadenas de caracteres que no satisface la condición de registro se registra o no en la BD de palabras relacionadas (etapa S55). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina si la relación de los conjuntos de cadenas de caracteres de 70 que tiene <palabra de búsqueda anterior - palabra de búsqueda> ya registrada disminuye en la BD 12d de conjuntos de cadenas de caracteres y se vuelve inferior a un umbral predeterminado (por ejemplo, 80%). Tal hecho se produce cuando otro conjunto de cadenas de caracteres que tiene la misma o similar palabra de búsqueda anterior aumenta.

En el caso en el que la condición de registro no se satisface (SÍ, en la etapa S55), el servidor 10 de registro de palabras relacionadas elimina el conjunto de cadenas de caracteres que no satisface la condición de registro de la BD de palabras relacionadas (etapa S56). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas elimina el conjunto de cadenas de caracteres que no satisface la condición de registro en los conjuntos de cadenas de caracteres que están ya registrados de la BD 12c de palabras relacionadas.

Después de la etapa S54, el servidor 10 de registro de palabras relacionadas determina si la extracción se realiza en todos los conjuntos de cadenas de caracteres o no (etapa S57). Específicamente, la unidad 14 de control del sistema del servidor 10 de registro de palabras relacionadas determina si el proceso de extracción de conjuntos de cadenas de caracteres en la etapa S51 se ha realizado o no sobre todas las palabras anteriores en el grupo de conjuntos de cadenas de caracteres en la BD 12d de conjuntos de cadenas de caracteres. En el caso en el que el proceso de extracción no se ha realizado sobre todos los conjuntos de cadenas de caracteres (NO en la etapa S57), la unidad 14 de control del sistema regresa a la etapa S11. En el caso en el que el proceso de extracción se ha realizado sobre todos los conjuntos de cadenas de caracteres (SÍ, en la etapa S57), el proceso de registro de palabras relacionadas finaliza.

De acuerdo con el modo de realización anterior, se recibe una consulta de búsqueda de una palabra de búsqueda introducida por el usuario. Las consultas de búsqueda anteriores obtenidas antes que la consulta de búsqueda recibida se extraen del registro de consultas de búsqueda en base a una condición de extracción de consultas de búsqueda predeterminada. La palabra de búsqueda anterior de la consulta de búsqueda anterior extraída y la palabra de búsqueda de la consulta de búsqueda recibida se almacenan como un conjunto de cadenas de caracteres en la BD 12d de conjuntos de cadenas de caracteres. Cuando se satisface la condición de inicio de extracción de conjuntos de cadenas de caracteres predeterminada, se extrae el conjunto de cadenas de caracteres que tiene la palabra de búsqueda que es igual o similar a la palabra de búsqueda anterior. El conjunto de cadenas de caracteres que satisface la condición de registro predeterminada se registra como palabras relacionadas en la BD 12c de palabras relacionadas. De tal manera, una cantidad predeterminada de conjuntos de cadenas de caracteres hecho cada uno por la consulta de búsqueda recibida y la consulta de búsqueda anterior en base a la condición de extracción de consultas de búsqueda se almacenan bajo la condición de inicio de extracción de conjuntos de cadenas de caracteres, y el conjunto de cadenas de caracteres que satisface la condición de registro en los conjuntos de cadenas de caracteres almacenados se registra como las palabras relacionadas en la BD 12c de palabras relacionadas. Por lo tanto, la precisión de las palabras relacionadas puede mejorarse.

En el caso en el que el servidor 10 de registro de palabras relacionadas también almacena la información de identificación del usuario y una consulta de búsqueda anterior cuya información de identificación de usuario corresponde a la del usuario que introdujo la consulta de búsqueda se extrae como la condición de extracción de consultas de búsqueda, puede generarse un conjunto de cadenas de caracteres por el mismo usuario, de manera que la precisión de las palabras relacionadas puede mejorarse adicionalmente.

En el caso en el que el servidor 10 de registro de palabras relacionadas extrae una consulta de búsqueda anterior recibida dentro de un tiempo predeterminado desde la hora de recepción de la consulta de búsqueda como la condición de extracción de consultas de búsqueda, dado que puede extraerse la consulta de búsqueda anterior que se recibe dentro de un tiempo predeterminado y tiene una alta relatividad, la precisión de las palabras relacionadas puede mejorarse adicionalmente.

En el caso en el que el servidor 10 de registro de palabras relacionadas almacena adicionalmente información de categorías, y una consulta de búsqueda anterior se extrae en base a la información de categorías como la condición de extracción de consultas de búsqueda, pueden formarse palabras de búsqueda cuyas categorías son las mismas o relacionadas en un conjunto de cadenas de caracteres, de manera que la precisión de las palabras relacionadas puede mejorarse.

Cuando el servidor 10 de registro de palabras relacionadas extrae un conjunto de cadenas de caracteres en el caso en el que el número de elementos en un grupo de conjuntos de cadenas de caracteres excede un umbral predeterminado como la condición de inicio de extracción de conjuntos de cadenas de caracteres, pueden almacenarse conjuntos de cadenas de caracteres en una cierta escala. Dado que puede extraerse un conjunto de cadenas de caracteres apropiado de los conjuntos almacenados, la precisión de las palabras relacionadas puede

mejorarse adicionalmente.

5 Cuando el servidor 10 de registro de palabras relacionadas registra un conjunto de cadenas de caracteres como palabras relacionadas en el caso en el que el número de conjuntos de cadenas de caracteres que tienen la misma o similar palabra de búsqueda excede un umbral predeterminado en los conjuntos de cadenas de caracteres que tienen la misma o similar palabra de búsqueda anterior como la condición de registro, puede seleccionarse el conjunto de cadenas de caracteres de alta precisión, y la precisión de las palabras relacionadas puede mejorarse adicionalmente.

10 Cuando el servidor 10 de registro de palabras relacionadas registra un conjunto de cadenas de caracteres como palabras relacionadas en el caso en el que la relación de conjuntos de cadenas de caracteres que tienen la misma o similar palabra de búsqueda excede un umbral predeterminado en los conjuntos de cadenas de caracteres que tienen la misma o similar palabra de búsqueda anterior como la condición de registro, puede seleccionarse el conjunto de cadenas de caracteres de alta precisión, y la precisión de las palabras relacionadas puede mejorarse adicionalmente.

15 Además, en el caso en el que el número de consultas de búsqueda en la BD 12a de registro de consultas de búsqueda excede un umbral predeterminado como la condición de inicio de extracción de conjuntos de cadenas de caracteres en la etapa S51, el servidor 10 de registro de palabras relacionadas puede extraer un conjunto de cadenas de caracteres. En este caso, pueden almacenarse conjuntos de cadenas de caracteres en una cierta escala. Dado que puede extraerse un conjunto de cadenas de caracteres apropiado de los conjuntos almacenados, la precisión de las palabras relacionadas puede mejorarse adicionalmente.

20 Cuando ha pasado un tiempo predeterminado desde que se satisface la condición de inicio de extracción de conjuntos de cadenas de caracteres en la etapa S50, el servidor 10 de registro de palabras relacionadas puede extraer un conjunto de cadenas de caracteres. En este caso, pueden almacenarse conjuntos de cadenas de caracteres en una cierta escala. Dado que puede extraerse un conjunto de cadenas de caracteres apropiado de los conjuntos almacenados, la precisión de las palabras relacionadas puede mejorarse adicionalmente.

25 Además, pueden registrarse palabras relacionadas por una combinación de las operaciones de la implementación a modo de ejemplo y el modo de realización dado a conocer en el sistema de registro de palabras relacionadas. En este caso, puede construirse una base de datos de palabras relacionadas en la BD 12c de palabras relacionadas de manera que pueden distinguirse entre sí las palabras relacionadas en las que la distancia entre las cadenas de caracteres se considera como la operación de la implementación a modo de ejemplo y las palabras relacionadas en las que el orden de uso de consultas de búsqueda se considera como en la operación del modo de realización.

30 Como un procedimiento de utilización de la base de datos de palabras relacionadas, por ejemplo, en el caso en el que se introduce "xx" para una búsqueda en el terminal 30, el servidor 20 de suministro de información se refiere a una base de datos de palabras relacionadas construida por la operación en la implementación a modo de ejemplo en la BD 12c de palabras relacionadas y muestra "¿tal vez 'xxx'?" en la página web 40 como una palabra relacionada como se ilustra en la figura 6. A continuación, en el caso en el que se selecciona "xxx" y se realiza una búsqueda en el terminal 30, el servidor 20 de suministro de información puede hacer referencia a una base de datos de palabras relacionadas construida por la operación en el modo de realización en la BD 12c de palabras relacionadas y se muestra "yy" como una palabra relacionada en la página web 40.

45 En lugar de mostrar las palabras relacionadas paso a paso, estas palabras relacionadas se pueden mostrar al mismo tiempo.

EXPLICACIÓN DE NÚMEROS DE REFERENCIA

- 50 1: sistema de registro de palabras relacionadas
- 10: servidor de registro de palabras relacionadas (dispositivo de registro de palabras relacionadas)
- 55 12a: BD de registro de consulta de búsqueda
- 12b: BD de candidatos de palabras relacionadas
- 12c: BD de palabras relacionadas
- 60 12d: BD de conjuntos de cadenas de caracteres
- 20: servidor de suministro de información (dispositivo de procesamiento de información)
- 65 22a: BD de productos

30: terminal

REIVINDICACIONES

1. Un dispositivo (10) de registro de palabras relacionadas que comprende:

un medio de recepción configurado para recibir una consulta de búsqueda de una palabra de búsqueda introducida por el usuario;

un medio (12a) de almacenamiento de consultas de búsqueda configurado para almacenar las consultas de búsqueda recibidas según el orden de recepción;

un medio de extracción de consultas de búsqueda configurado para extraer, del medio (12a) de almacenamiento de consultas de búsqueda, una consulta de búsqueda anterior cuyo orden de recepción es anterior al de la consulta de búsqueda recibida en base a una condición de extracción de consultas de búsqueda predeterminada, en el que el medio de extracción de consultas de búsqueda está configurado para extraer una consulta de búsqueda anterior recibida en un tiempo predeterminado desde el momento de recepción de la consulta de búsqueda como la condición de extracción de consulta de búsqueda;

un medio (12d) de almacenamiento de grupos de cadenas de caracteres configurado para almacenar, como un conjunto (70) de cadenas de caracteres, una palabra de búsqueda anterior de la consulta de búsqueda anterior extraída y una palabra de búsqueda de la consulta de búsqueda recibida;

un medio de extracción de cadenas de caracteres configurado para extraer conjuntos (70) de cadenas de caracteres que tienen las palabras de búsqueda anteriores iguales o similares del medio (12d) de almacenamiento de conjuntos de cadenas de caracteres de acuerdo con una condición de inicio de extracción de conjuntos de cadenas de caracteres predeterminada, en el que el medio de extracción de cadenas de caracteres está configurado para extraer el conjunto de cadenas de caracteres en el caso en el que el número de consultas de búsqueda en el medio de almacenamiento de consultas de búsqueda o el número total de grupos de conjuntos de cadenas de caracteres exceda un umbral predeterminado como una condición de inicio de extracción de conjuntos de cadenas de caracteres;

un medio de especificación de palabras relacionadas configurado para especificar un conjunto de cadenas de caracteres como palabras relacionadas del conjunto de cadenas de caracteres extraído en base a una condición de registro predeterminada, en el que la condición de registro predeterminada se satisface en el caso en el que el número o la proporción de conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda entre los conjuntos de cadenas de caracteres extraídos exceda un umbral predeterminado; y

un medio de registro de palabras relacionadas configurado para registrar el conjunto de cadenas de caracteres especificado como palabras relacionadas en una base (12c) de datos de palabras relacionadas, en el que las palabras de búsqueda, respectivamente palabras de búsqueda anteriores, son similares entre sí cuando el valor de similitud en base a la distancia entre sus cadenas de caracteres es igual o mayor que un umbral predeterminado.

2. El dispositivo de registro de palabras relacionadas de acuerdo con la reivindicación 1, en el que el medio (12a) de almacenamiento de consultas de búsqueda está configurado además para almacenar información de identificación de usuario, y el medio de extracción de consultas de búsqueda está configurado para extraer la consulta de búsqueda anterior cuya información de identificación de usuario coincide con la del usuario que introdujo la consulta de búsqueda como la condición de extracción de consulta de búsqueda.

3. El dispositivo de registro de palabras relacionadas de acuerdo con una cualquiera de las reivindicaciones 1 a 2, en el que el medio de almacenamiento de consultas de búsqueda está configurado además para almacenar información de categorías, y

el medio de extracción de consultas de búsqueda está configurado para extraer la consulta de búsqueda anterior en base a la información de categorías como la condición de extracción de consulta de búsqueda.

4. El dispositivo de registro de palabras relacionadas de acuerdo con la reivindicación 1, en el que el medio de extracción de cadenas de caracteres está configurado para extraer el conjunto de cadenas de caracteres en el caso en el que transcurre el tiempo predeterminado dado que se satisface la condición de inicio de extracción del conjunto de cadenas de caracteres.

5. Un sistema (5) de servidor que comprende el dispositivo (10) de registro de palabras relacionadas de acuerdo con una cualquiera de las reivindicaciones 1 a 4 y un dispositivo (20) de procesamiento de información para procesar la información con referencia a una base (12c) de datos de palabras relacionadas del dispositivo de registro de palabras relacionadas, que comprende:

un medio de extracción de palabras relacionadas configurado para extraer una palabra relacionada

correspondiente a una palabra de búsqueda de la consulta de búsqueda recibida con referencia a la base de datos de palabras relacionadas; y

5 un medio de salida de palabras relacionadas configurado para transmitir la palabra relacionada extraída por el medio de extracción de palabras relacionadas.

6. Un procedimiento de registro de palabras relacionadas en un dispositivo de registro de palabras relacionadas para registrar una palabra relacionada, que comprende:

10 una etapa de recepción (S41) de recibir una consulta de búsqueda de una palabra de búsqueda introducida por el usuario;

15 una etapa de almacenamiento de consultas de búsqueda (S45) de almacenar las consultas de búsqueda recibidas según el orden de recepción en un medio de almacenamiento de consultas de búsqueda;

20 una etapa de extracción de consultas de búsqueda (S42, S43) de extraer, del medio de almacenamiento de consultas de búsqueda, una consulta de búsqueda anterior cuyo orden de recepción es anterior al de la consulta de búsqueda recibida en base a una condición de extracción de consultas de búsqueda predeterminada, en el que la etapa de extracción de consultas de búsqueda extrae una consulta de búsqueda anterior recibida dentro de un tiempo predeterminado desde la hora de recepción de la consulta de búsqueda como la condición de extracción de consulta de búsqueda;

25 una etapa de almacenamiento de conjuntos de cadenas de caracteres (S44) de almacenar, como un conjunto de cadenas de caracteres, una palabra de búsqueda anterior de la consulta de búsqueda anterior extraída y una palabra de búsqueda de la consulta de búsqueda recibida en un medio (12d) de almacenamiento de conjuntos de cadenas de caracteres;

30 una etapa de extracción de cadenas de caracteres (S51) de extraer conjuntos (70) de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda anteriores del medio de almacenamiento de conjuntos de cadenas de caracteres según una condición de inicio de extracción del conjunto de cadenas de caracteres predeterminada, en el que el medio de extracción de cadenas de caracteres está configurado para extraer el conjunto de cadenas de caracteres en el caso en el que el número de consultas de búsqueda en el medio de almacenamiento de consultas de búsqueda o el número total de grupos de conjuntos de cadenas de caracteres exceda un umbral predeterminado como una condición de inicio de extracción del conjunto de cadenas de caracteres;

35 una etapa de especificación de palabras relacionadas (S52, S53) de especificar un conjunto de cadenas de caracteres como palabras relacionadas de los conjuntos de cadenas de caracteres extraídos en base a una condición de registro predeterminada, en el que la condición de registro predeterminada se satisface en el caso en el que el número o la proporción de conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda entre los conjuntos de cadenas de caracteres extraídos excede un umbral predeterminado; y

40 una etapa de registro de palabras relacionadas (S54) de registrar el conjunto de cadenas de caracteres especificado como palabras relacionadas en una base de datos de palabras relacionadas, en el que palabras de búsqueda, respectivamente palabras de búsqueda anteriores, son similares entre sí cuando el valor de similitud en base a la distancia entre sus cadenas de caracteres es igual o mayor que un umbral predeterminado.

45 7. Un programa para un dispositivo de registro de palabras relacionadas, para hacer que un ordenador funcione como:

50 un medio de recepción que recibe una consulta de búsqueda de una palabra de búsqueda introducida por el usuario;

55 un medio (12a) de almacenamiento de consultas de búsqueda que almacena las consultas de búsqueda recibidas según el orden de recepción;

60 un medio de extracción de consultas de búsqueda que extrae, del medio de almacenamiento de consultas de búsqueda, una consulta de búsqueda anterior cuyo orden de recepción es anterior al de la consulta de búsqueda recibida en base a una condición de extracción de consultas de búsqueda predeterminada, en el que el medio de extracción de consultas de búsqueda está configurado para extraer una consulta de búsqueda anterior recibida dentro de un tiempo predeterminado desde la hora de recepción de la consulta de búsqueda como la condición de extracción de consulta de búsqueda;

65 un medio (12d) de almacenamiento de conjuntos de cadenas de caracteres que almacena, como un conjunto (70) de cadenas de caracteres, una palabra de búsqueda anterior de la consulta de búsqueda anterior extraída y una palabra de búsqueda de la consulta de búsqueda recibida;

un medio de extracción de cadenas de caracteres que extrae conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda anteriores del medio de almacenamiento de conjuntos de cadenas de caracteres según una condición de inicio de extracción del conjunto de cadenas de caracteres predeterminada, en el que el medio de extracción de cadenas de caracteres está configurado para extraer el conjunto de cadenas de caracteres en el caso en el que el número de consultas de búsqueda en el medio de almacenamiento de consultas de búsqueda o el número total de grupos de conjuntos de cadenas de caracteres exceda un umbral predeterminado como una condición de inicio de extracción del conjunto de cadenas de caracteres;

un medio de especificación de palabras relacionadas que especifica un conjunto de cadenas de caracteres como palabras relacionadas de los conjuntos de cadenas de caracteres extraídos en base a una condición de registro predeterminada, en el que la condición de registro predeterminada se satisface en el caso en el que el número o la proporción de conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda entre los conjuntos de cadenas de caracteres extraídos excede un umbral predeterminado; y

un medio de registro de palabras relacionadas que registra el conjunto de cadenas de caracteres especificado como palabras relacionadas en una base de datos de palabras relacionadas, en el que palabras de búsqueda, respectivamente palabras de búsqueda anteriores, son similares entre sí cuando el valor de similitud en base a la distancia entre sus cadenas de caracteres es igual o mayor que un umbral predeterminado.

8. Un medio de almacenamiento legible por ordenador que almacena un programa para un dispositivo de registro de palabras relacionadas, para hacer que un ordenador funcione como:

un medio de recepción que recibe una consulta de búsqueda de una palabra de búsqueda introducida por el usuario;

un medio (12a) de almacenamiento de consultas de búsqueda que almacena las consultas de búsqueda recibidas según el orden de recepción;

un medio de extracción de consultas de búsqueda que extrae, del medio de almacenamiento de consultas de búsqueda, una consulta de búsqueda anterior cuyo orden de recepción es anterior al de la consulta de búsqueda recibida en base a una condición de extracción de consultas de búsqueda predeterminada, en el que el medio de extracción de consultas de búsqueda está configurado para extraer una consulta de búsqueda anterior recibida dentro de un tiempo predeterminado desde la hora de recepción de la consulta de búsqueda como la condición de extracción de consulta de búsqueda;

un medio (12d) de almacenamiento de conjuntos de cadenas de caracteres que almacena, como un conjunto de cadenas de caracteres, una palabra de búsqueda anterior de la consulta de búsqueda anterior extraída y una palabra de búsqueda de la consulta de búsqueda recibida;

un medio de extracción de cadenas de caracteres que extrae conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda anteriores del medio de almacenamiento de conjuntos de cadenas de caracteres según una condición de inicio de extracción del conjunto de cadenas de caracteres predeterminada, en el que el medio de extracción de cadenas de caracteres está configurado para extraer el conjunto de cadenas de caracteres en el caso en el que el número de consultas de búsqueda en el medio de almacenamiento de consultas de búsqueda o el número total de grupos de conjuntos de cadenas de caracteres exceda un umbral predeterminado como una condición de inicio de extracción del conjunto de cadenas de caracteres;

un medio de especificación de palabras relacionadas que especifica un conjunto de cadenas de caracteres como palabras relacionadas de los conjuntos de cadenas de caracteres extraídos en base a una condición de registro predeterminada, en el que la condición de registro predeterminada se satisface en el caso en el que el número o la proporción de conjuntos de cadenas de caracteres que tienen las mismas o similares palabras de búsqueda entre los conjuntos de cadenas de caracteres extraídos excede un umbral predeterminado; y

un medio de registro de palabras relacionadas que registra el conjunto de cadenas de caracteres especificado como palabras relacionadas en una base de datos de palabras relacionadas, en el que palabras de búsqueda, respectivamente palabras de búsqueda anteriores, son similares entre sí cuando el valor de similitud en base a la distancia entre sus cadenas de caracteres es igual o mayor que un umbral predeterminado.

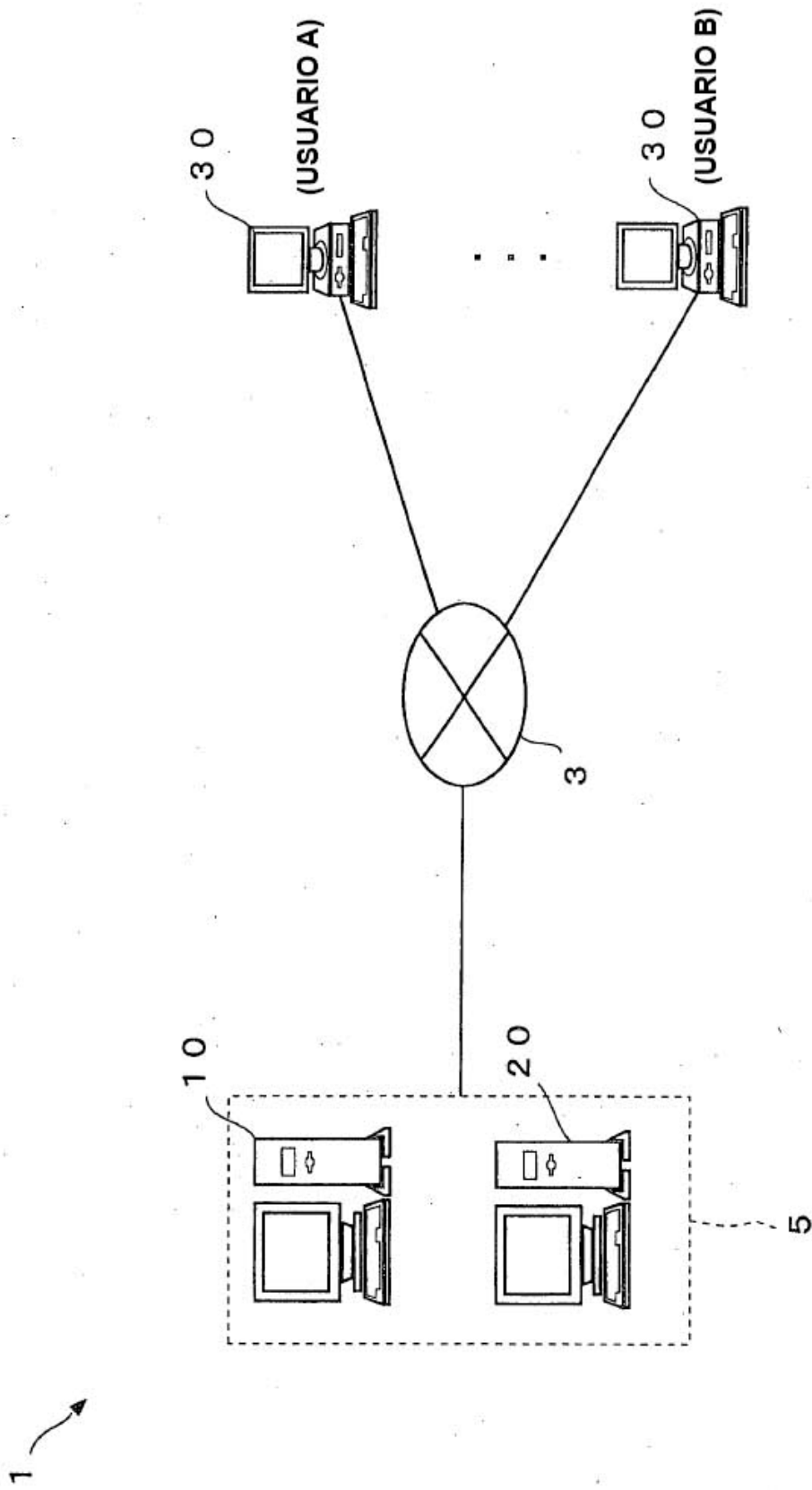


FIG.1

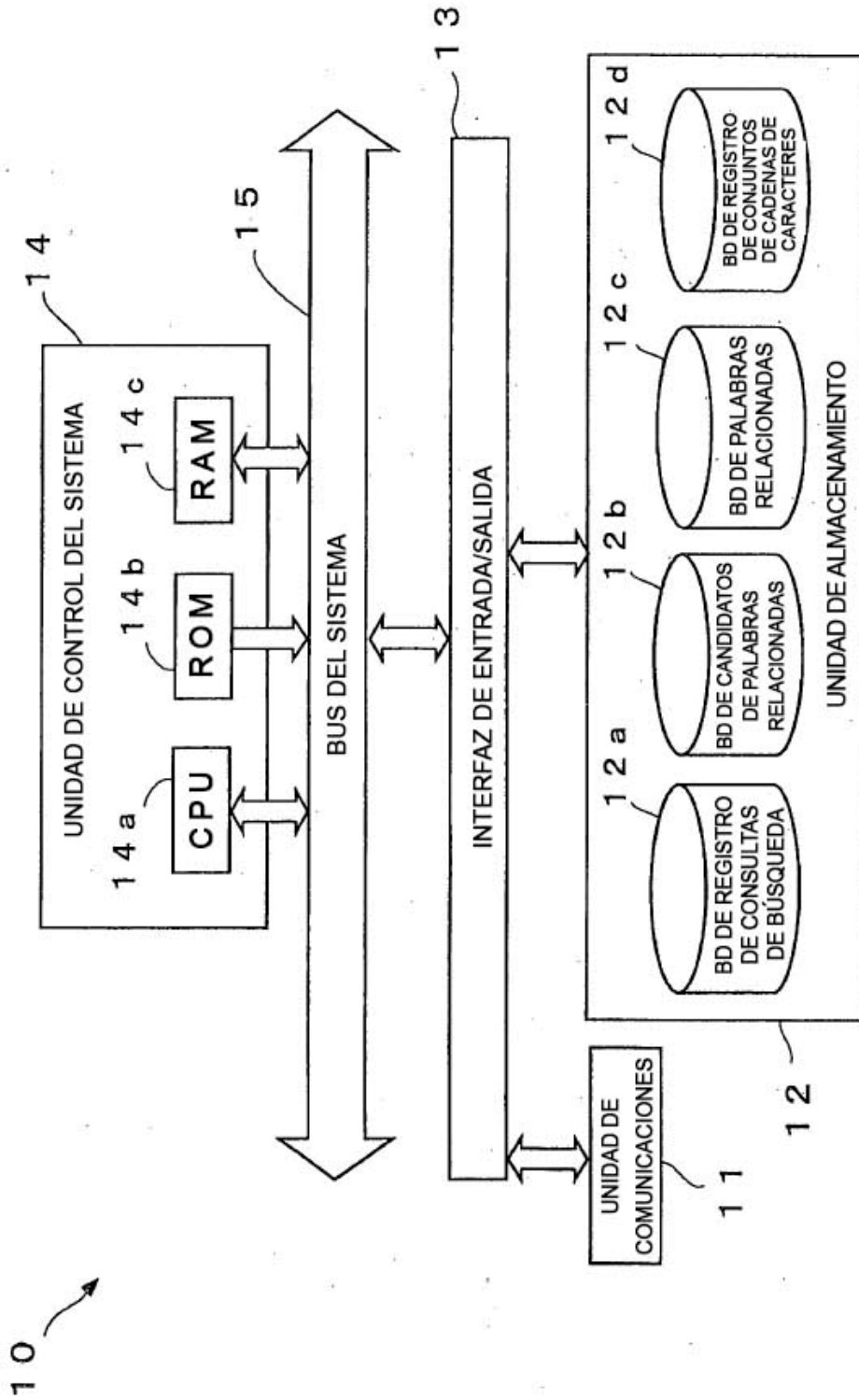


FIG.2

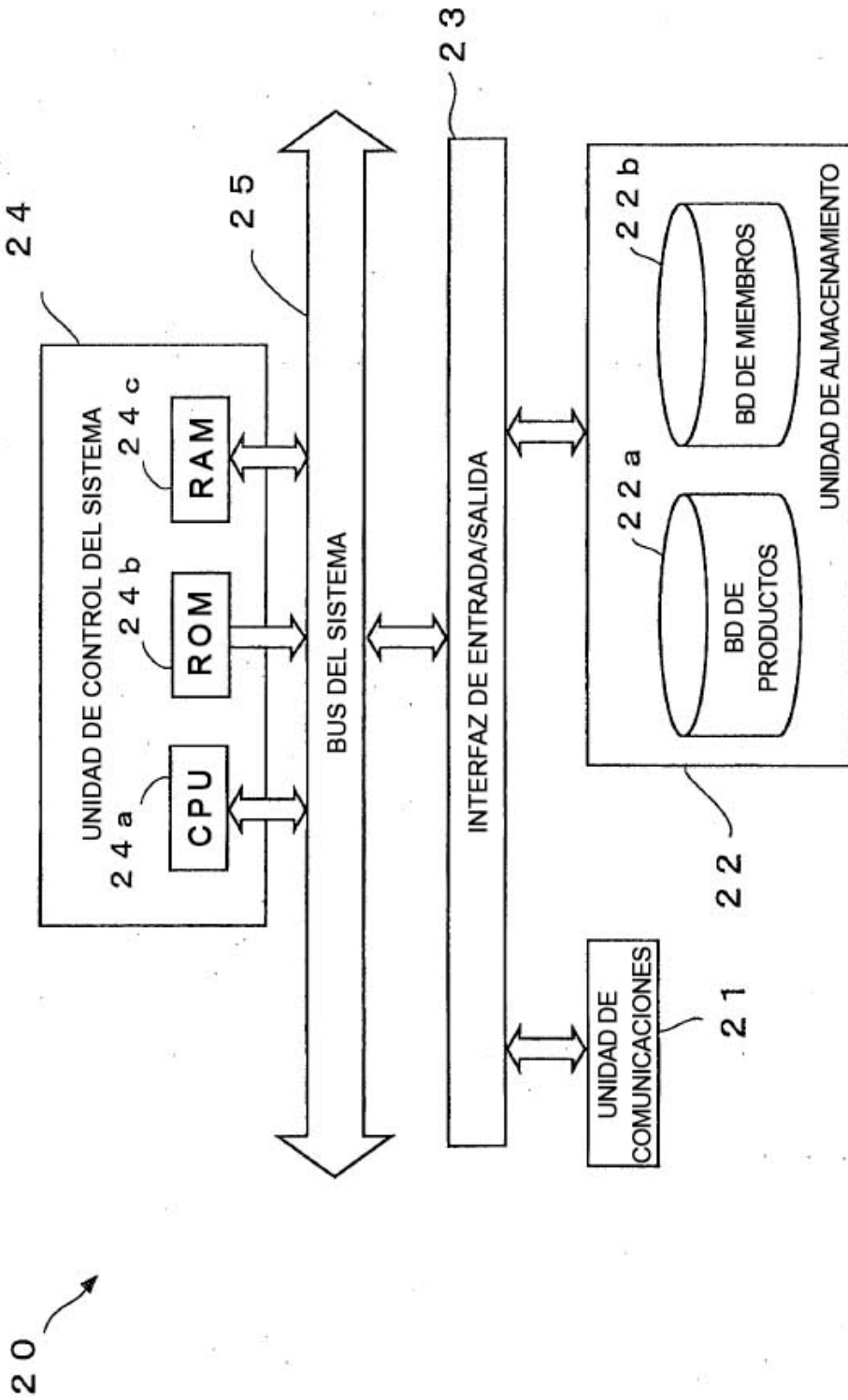


FIG.3

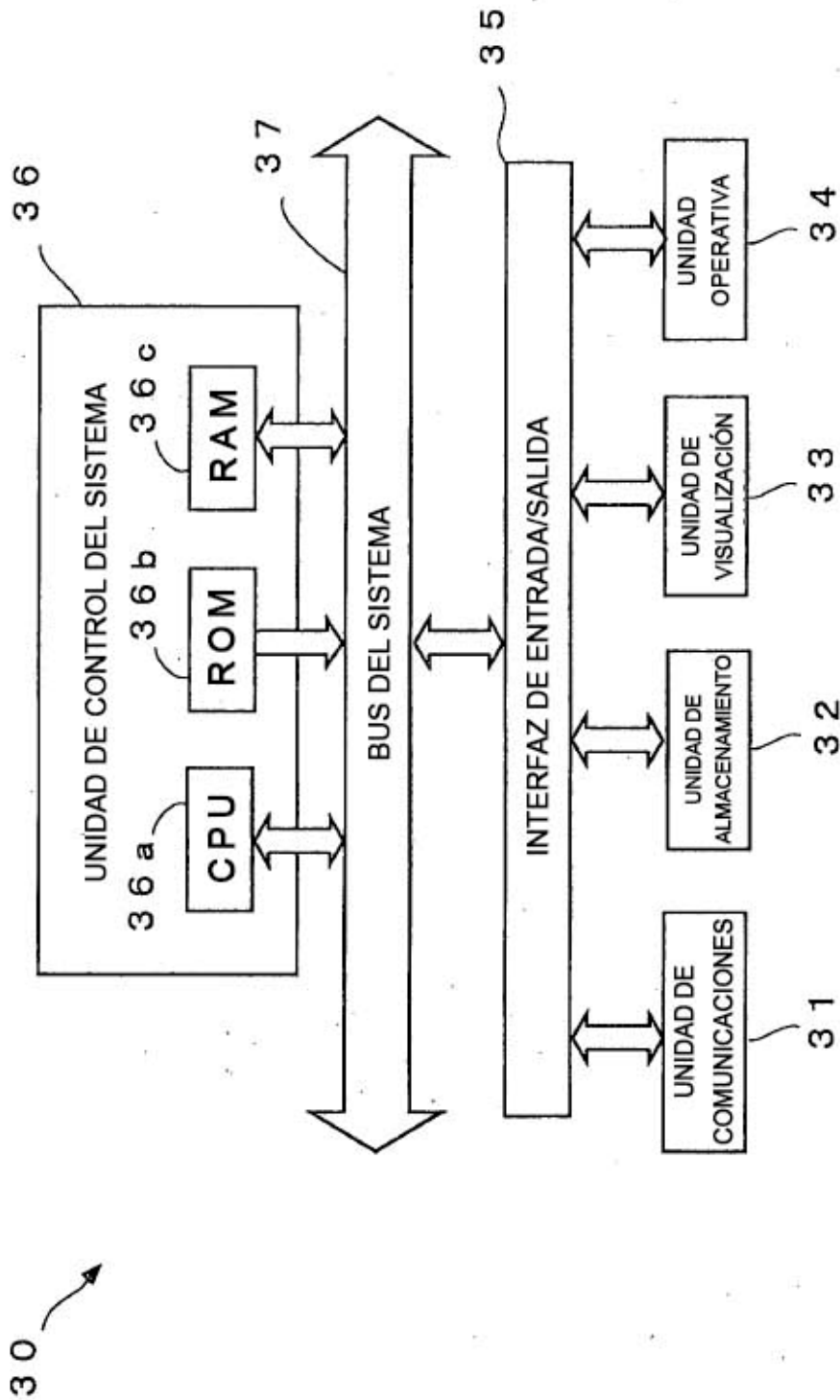


FIG.4

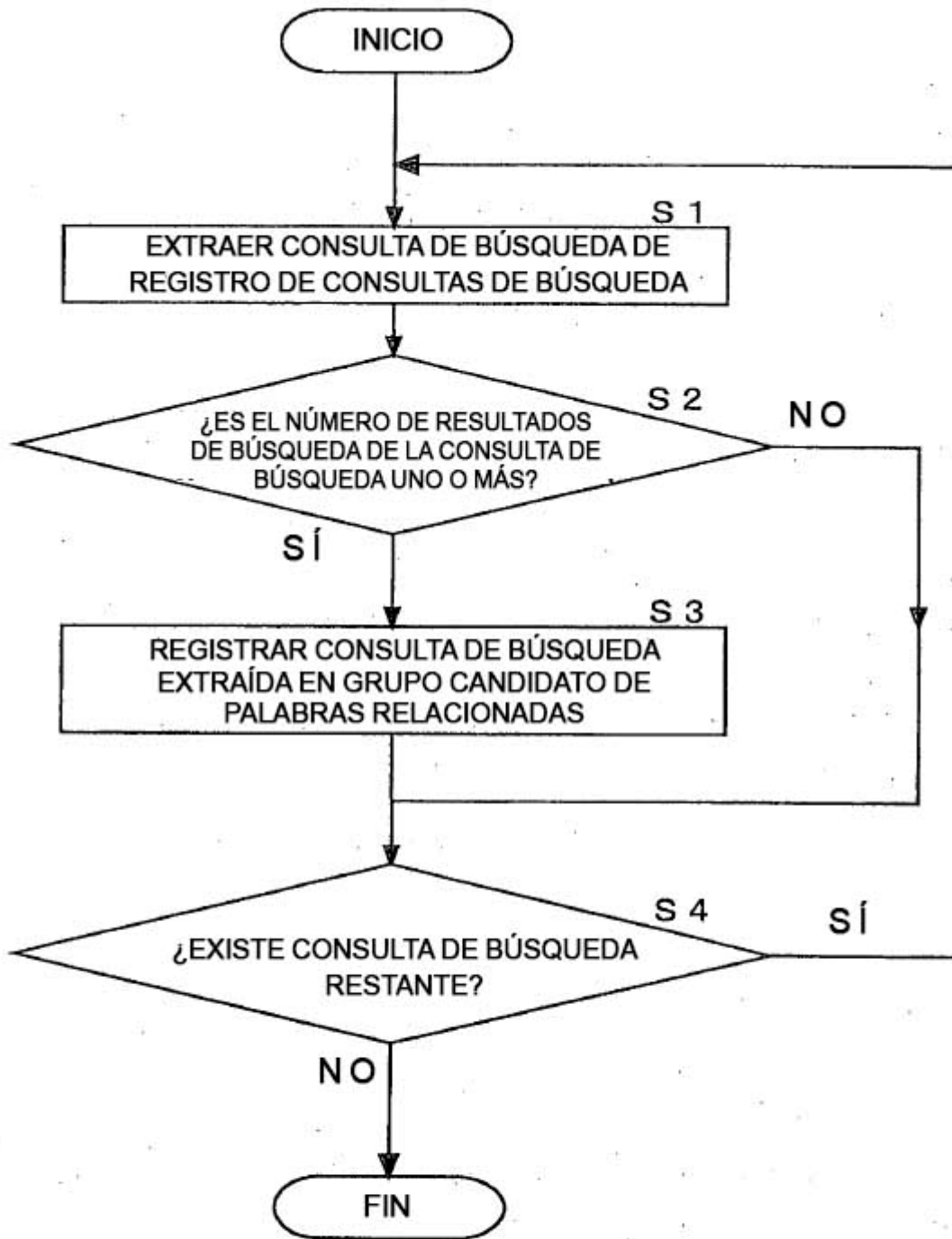


FIG.5

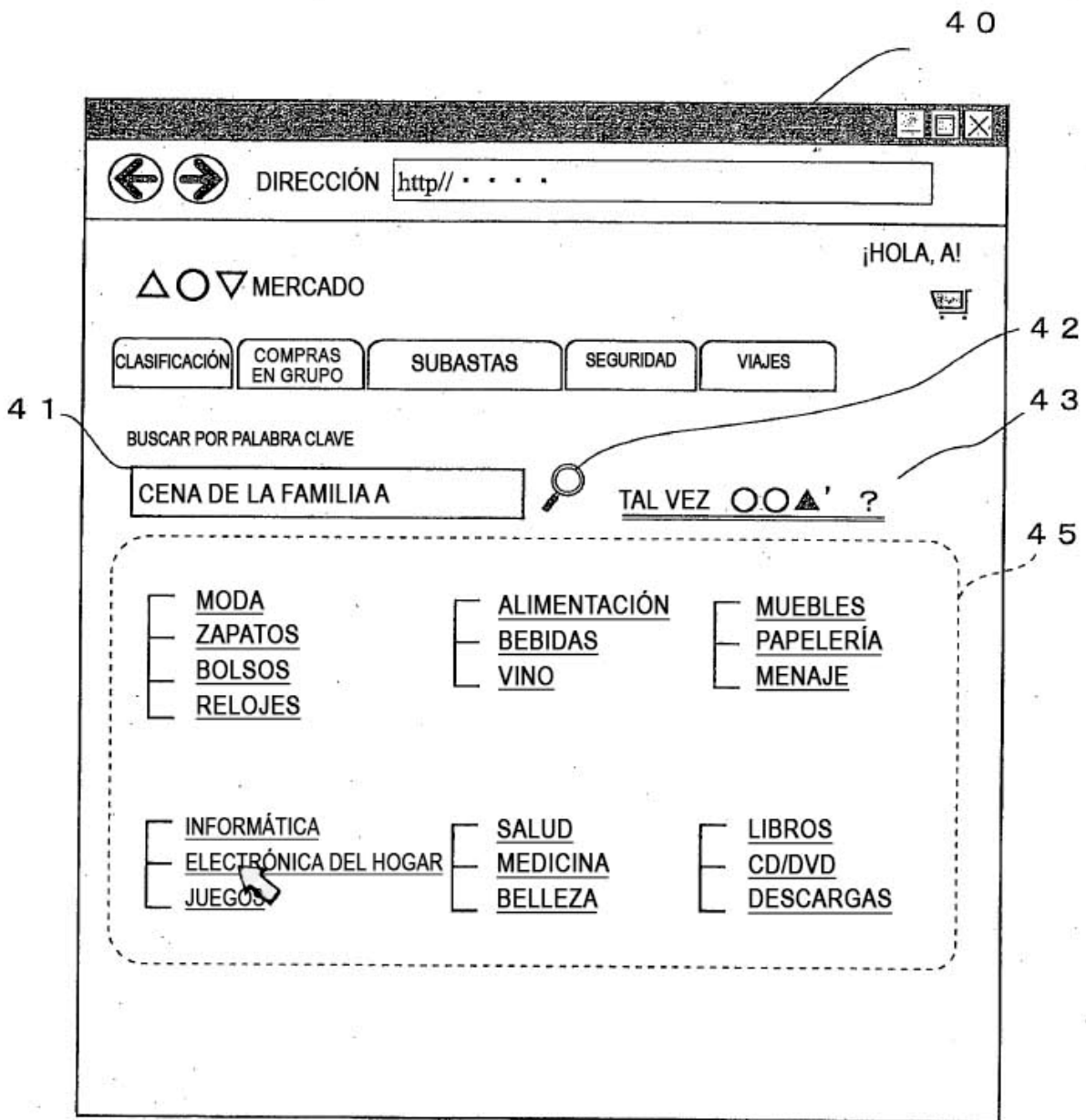


FIG.6

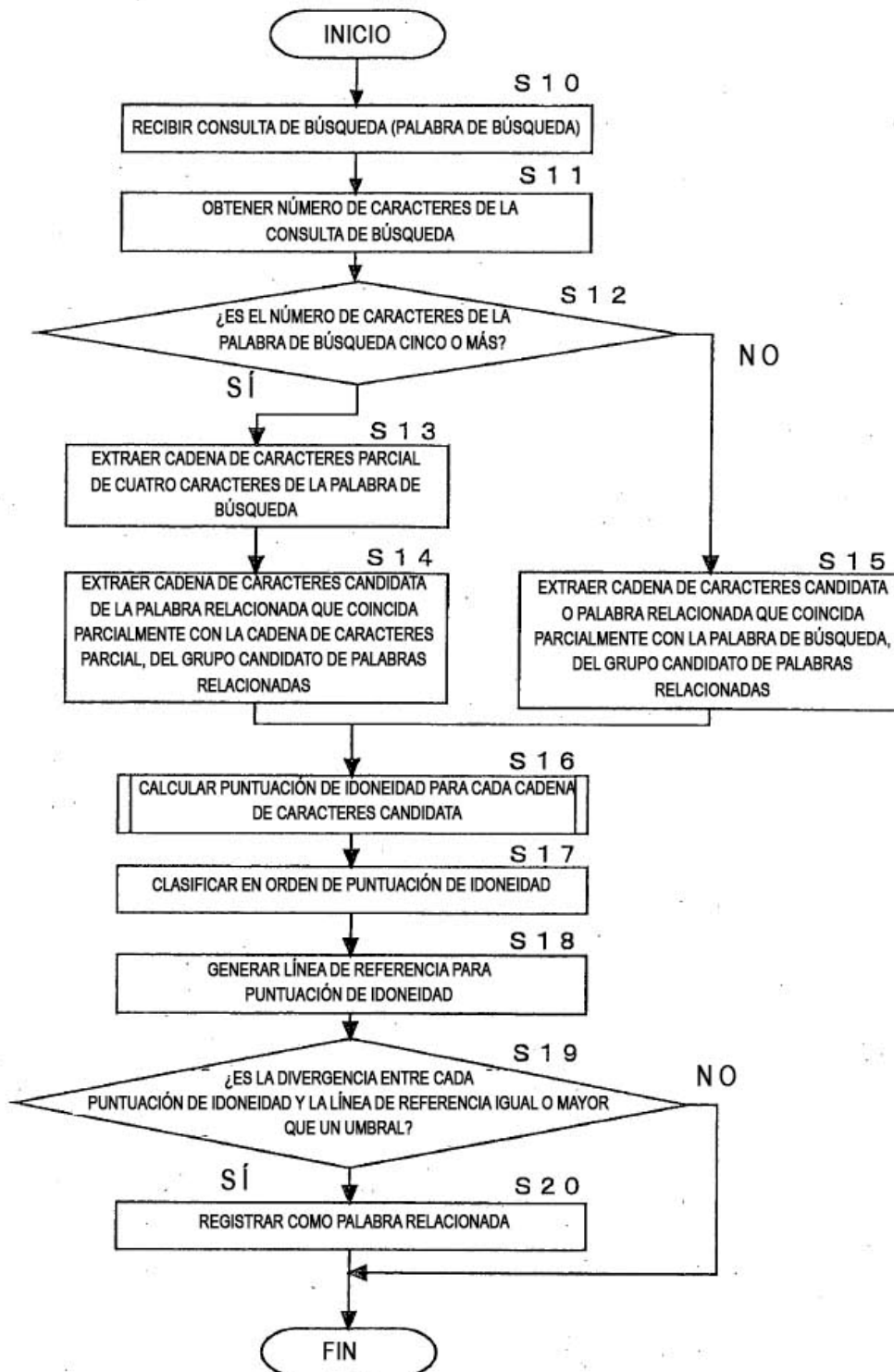
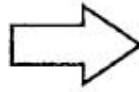
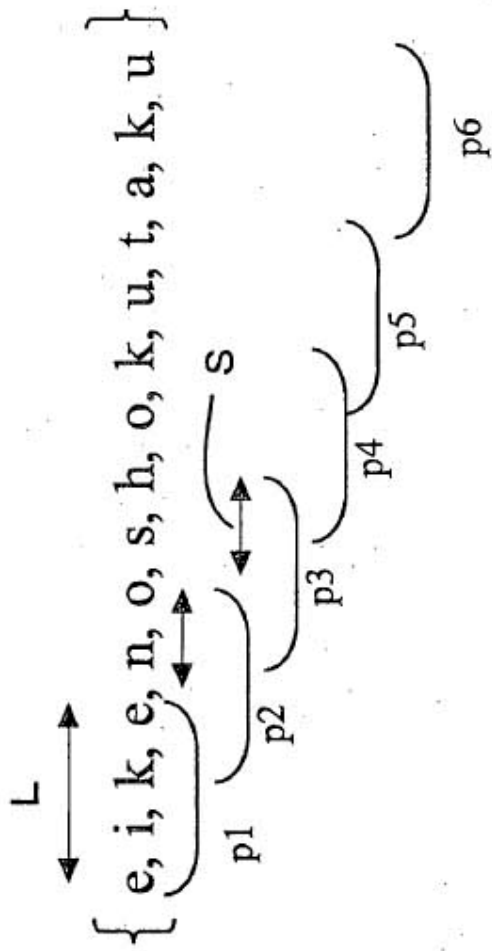


FIG.7



Consulta = "eike" O "keno" O "nosh" O "shok" O "okut" O "utak"

Below the words, six curly braces labeled p1 through p6 are positioned under the words: p1 under "eike", p2 under "keno", p3 under "nosh", p4 under "shok", p5 under "okut", and p6 under "utak".

Between the word "keno" and "nosh", between "nosh" and "shok", and between "shok" and "okut", there is a curly brace labeled "OR".

FIG. 8

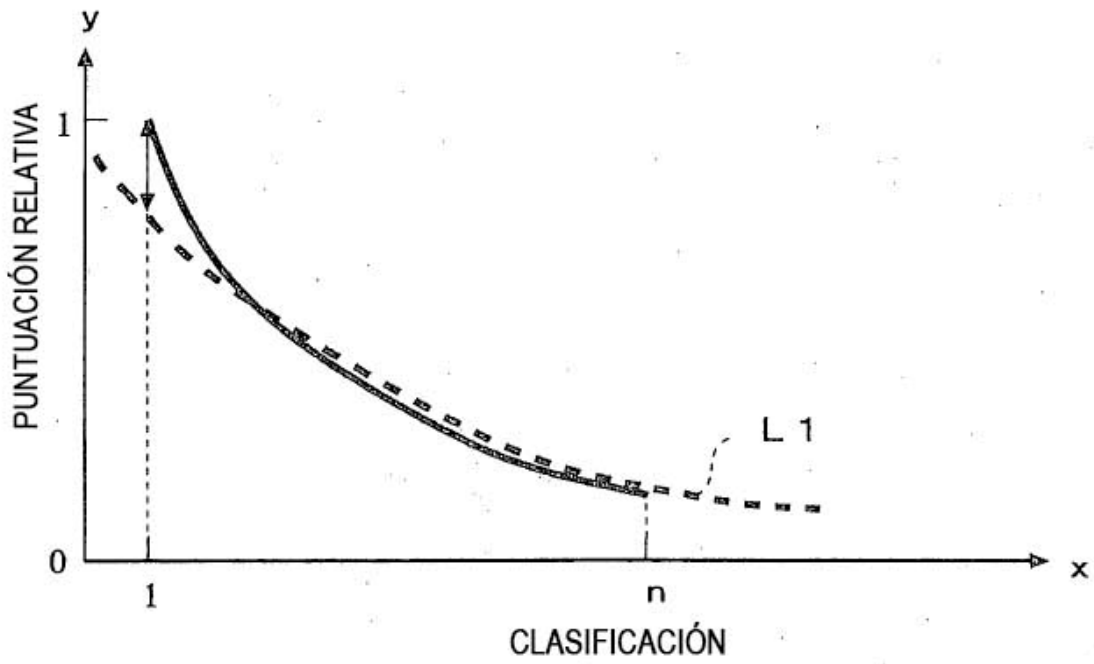


FIG. 9 A

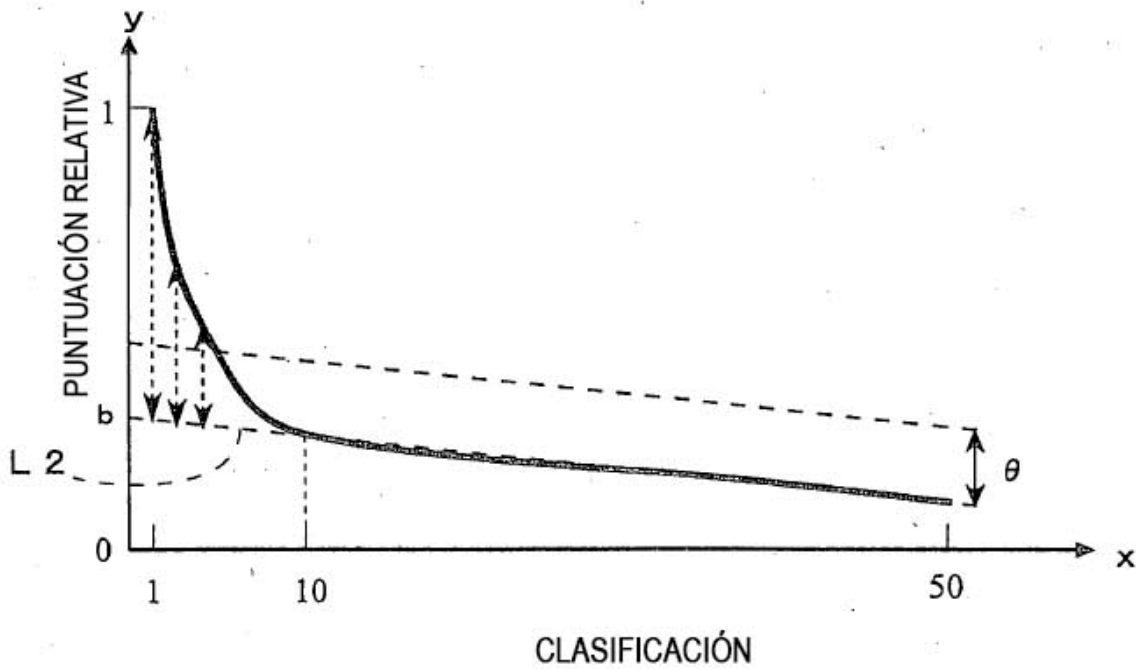


FIG. 9 B

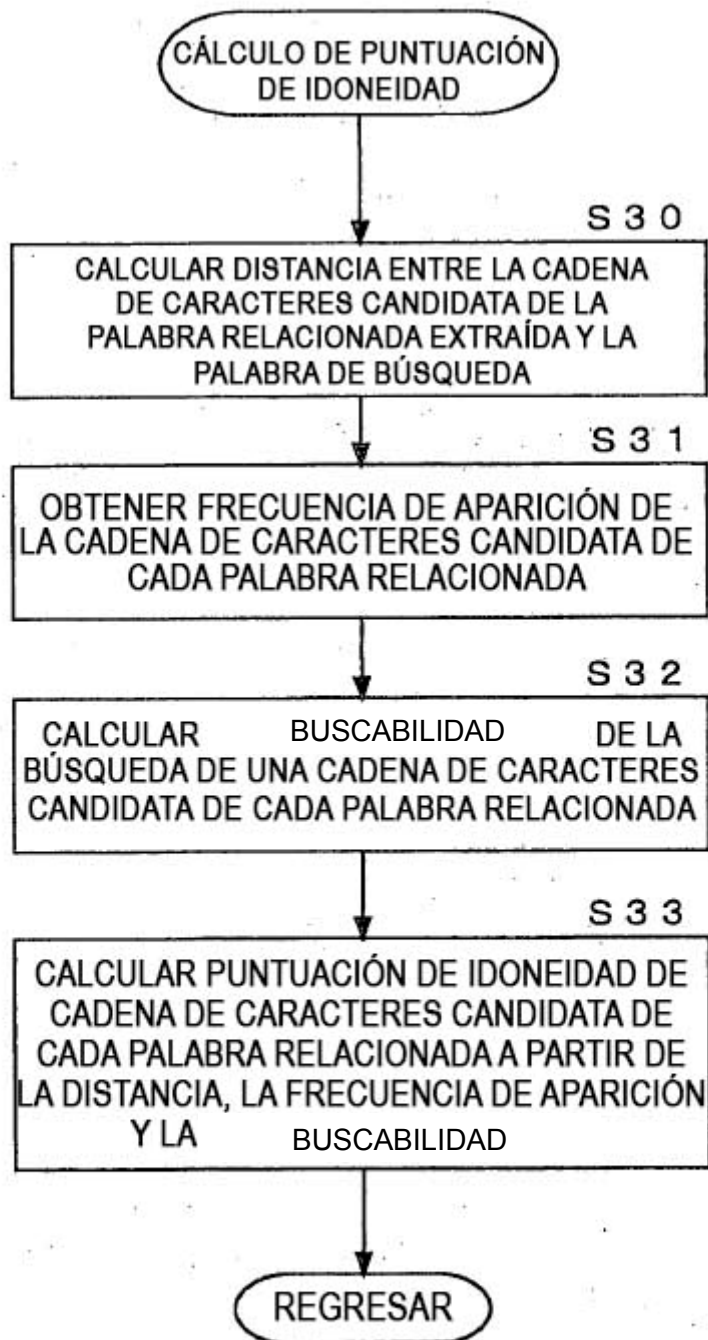


FIG.10

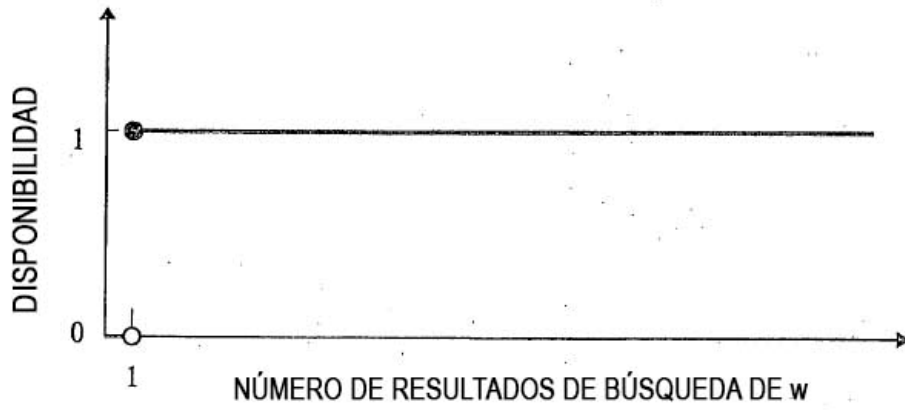


FIG.11 A



FIG.11 B



FIG.11 C

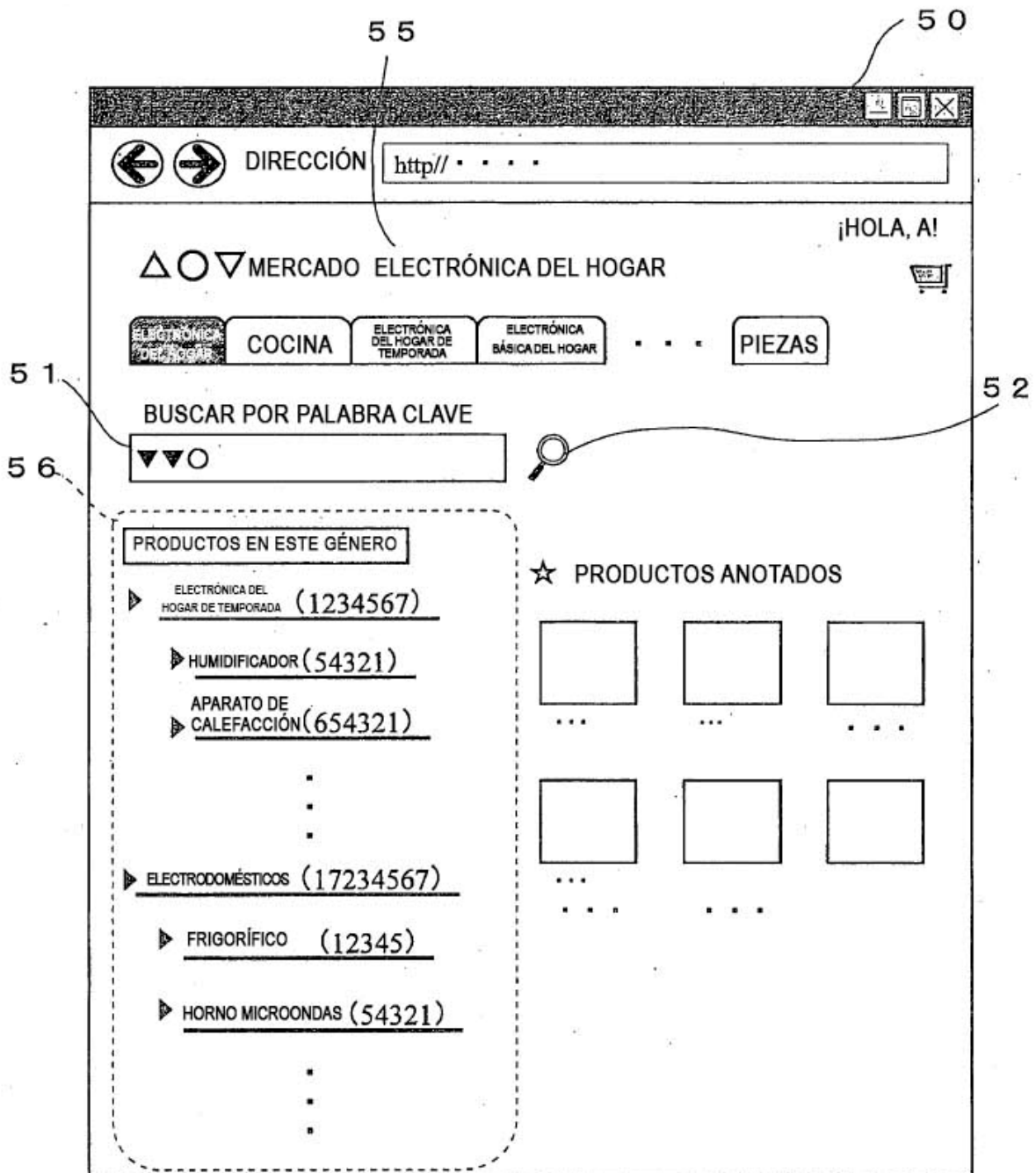


FIG.12

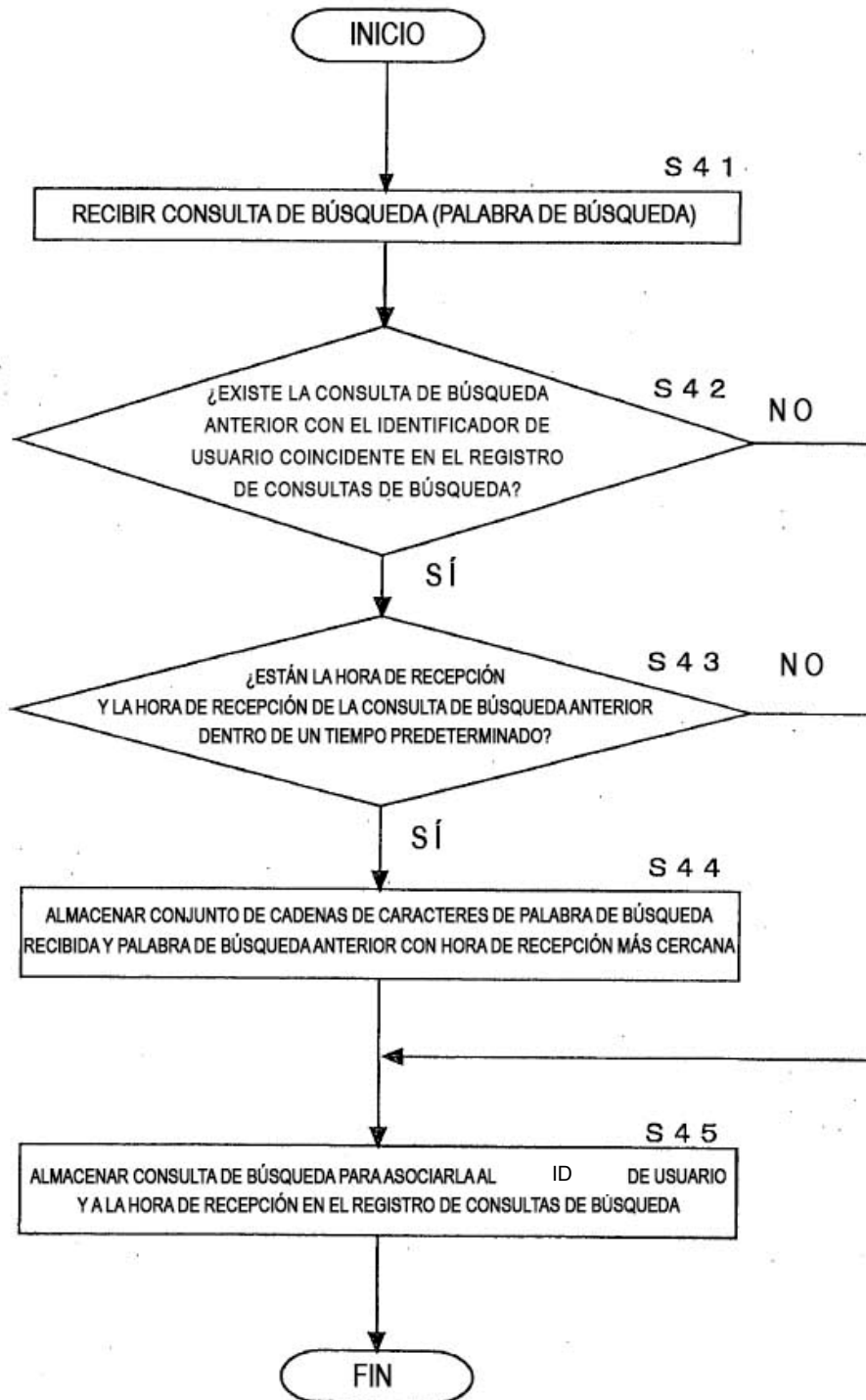


FIG.13

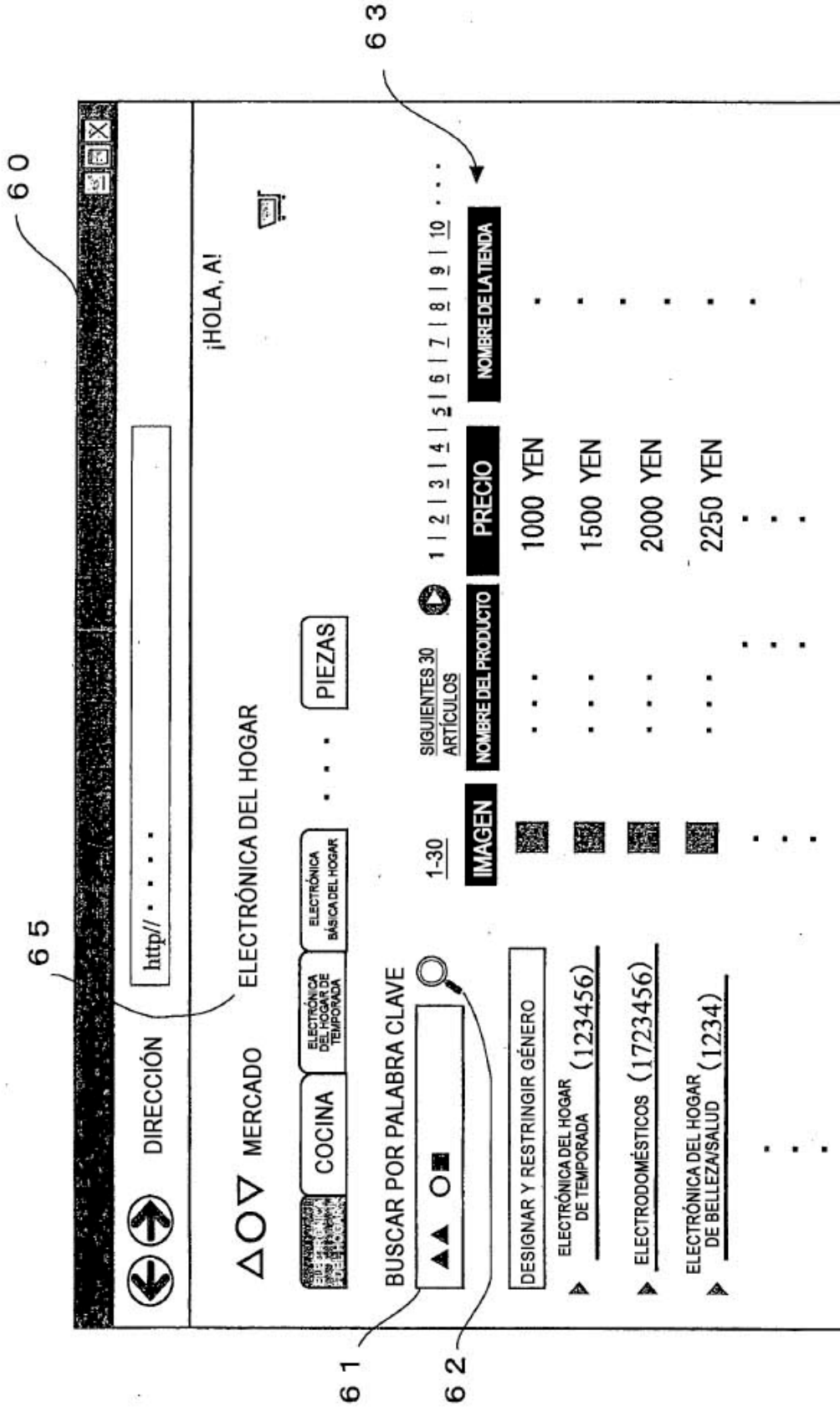


FIG.14

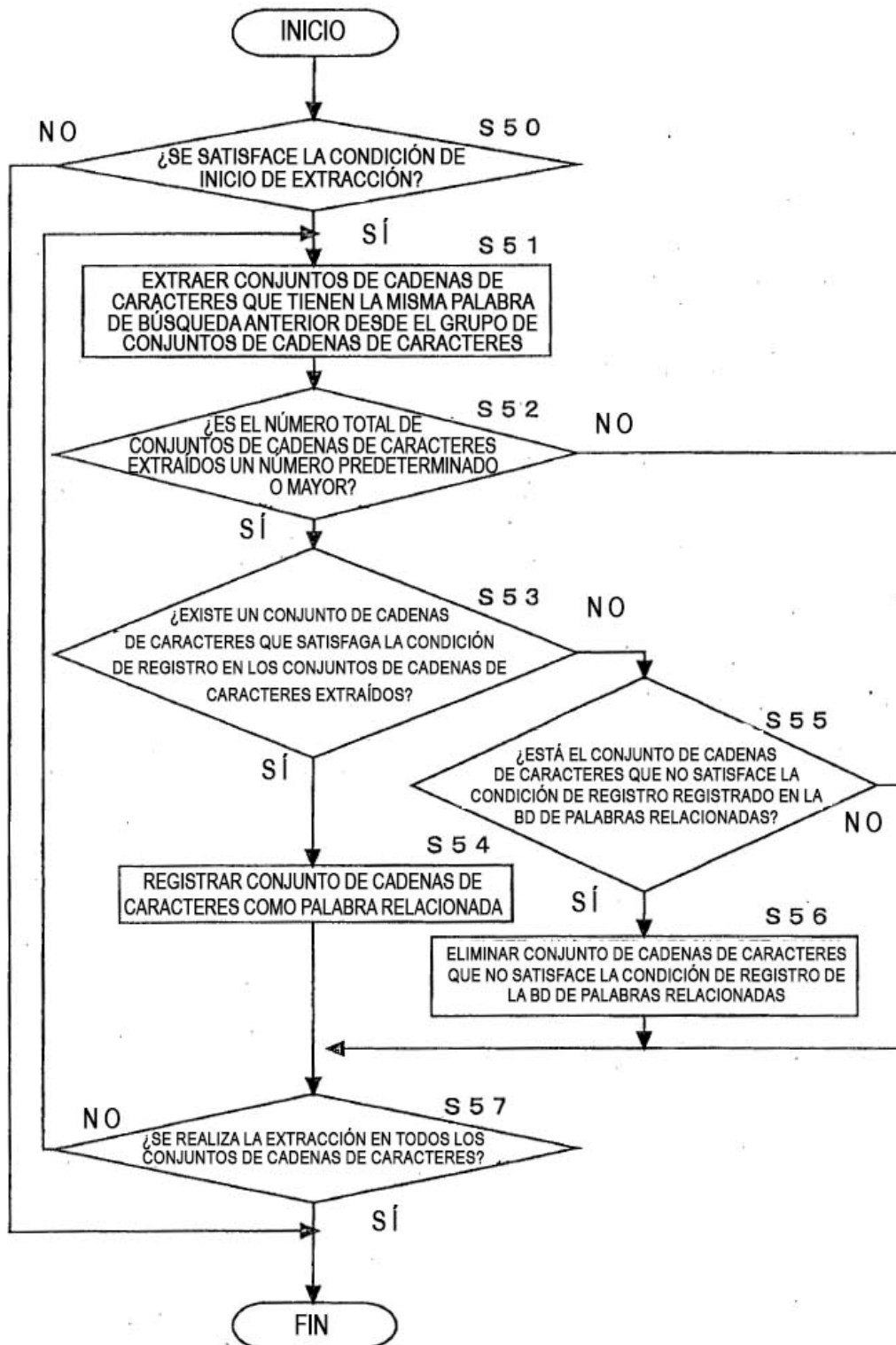


FIG15

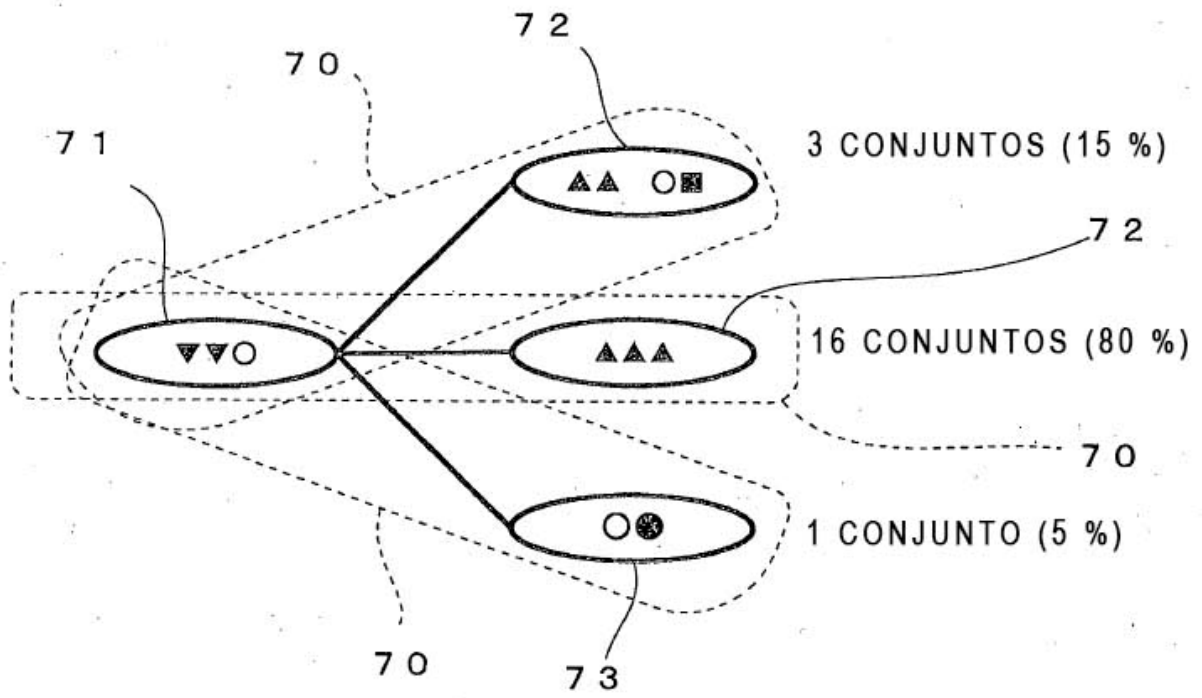


FIG.16