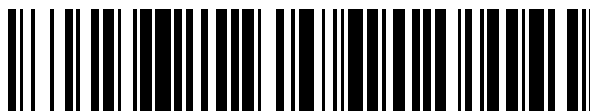


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 374 881**

51 Int. Cl.:  
**G06F 17/30** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **00940336 .1**  
96 Fecha de presentación: **07.06.2000**  
97 Número de publicación de la solicitud: **1389317**  
97 Fecha de publicación de la solicitud: **18.02.2004**

54 Título: **MÁQUINA DE BÚSQUEDA DE TEXTOS.**

30 Prioridad:  
**09.06.1999 DE 19926195**  
**13.12.1999 DE 19960081**

45 Fecha de publicación de la mención BOPI:  
**22.02.2012**

45 Fecha de la publicación del folleto de la patente:  
**22.02.2012**

73 Titular/es:  
**Berberich, Olaf**  
**Breiten Dyk 14**  
**47803 Krefeld, DE**

72 Inventor/es:  
**Berberich, Olaf**

74 Agente: **No consta**

ES 2 374 881 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

Máquina de búsqueda de textos

5 La presente invención se refiere a un procedimiento para la formación de una unión de una palabra a buscar constituida por caracteres lingüísticos muy pequeños (monemas) y un archivo, de manera que en una primera etapa del proceso, la palabra a buscar es predeterminada mediante una unidad de introducción, en especial, el teclado de un ordenador. La invención se refiere también a un sistema para la realización del procedimiento.

10 Estos procedimientos en los que el usuario puede buscar informaciones mediante una palabra de búsqueda o concepto de búsqueda predeterminado por el mismo, constituido mediante "monemas" en uno o varios archivos principalmente contenidos en bases de datos, son conocidos según múltiples formas de realización. Estos procedimientos son realizados preferentemente por programas de ordenador que están dotados de los correspondientes algoritmos de búsqueda para llevar a cabo la búsqueda y que son puestos a disposición por los  
15 gestionadores de las bases de datos o por proveedores independientes. Estos programas de búsqueda son ofertados también como las designadas "máquinas de búsqueda" en gran número para la búsqueda de artículos y/o servicios, en Internet.

20 Todos los programas de búsqueda conocidos hasta el momento funcionan según el mismo principio. De acuerdo con el mismo, la cantidad de datos existentes para la búsqueda son ordenados sistemáticamente, por ejemplo, según el alfabeto, en forma de un archivo de índice. Mediante esta ordenación, es sencillo buscar en los lugares adecuados, igual que en una guía telefónica, según el concepto deseado.

25 Es desfavorable en todos los programas de búsqueda conocidos que buscan de manera invariable de acuerdo con el concepto predeterminado y las desviaciones del concepto concreto son permisibles solamente dentro del ámbito de un llamado "comodín" predeterminado (Joker). De esta manera, encuentran exclusivamente una reproducción del concepto predeterminado o de la parte componente predeterminada. A los programas conocidos, no les es posible, por ejemplo, encontrar el plural de una palabra a buscar cuando se ha predeterminado solamente el singular. En especial, no es posible conseguir resultados satisfactorios con conceptos predeterminados erróneos,  
30 por ejemplo, conceptos escritos de manera errónea por descuido.

Por el documento WO 98/37474 A, se conoce un procedimiento según la parte introductoria de la reivindicación 1. Según dicho documento WO 98/37474 A, se facilitan frases ("keynotes") para buscar de estas frases palabras individuales ("keywords") ("palabras clave"). En este caso, la función indicada no reconocería una palabra a buscar  
35 cuando se hubiera escrito erróneamente en un lugar que no influye en el sentido.

Es objetivo de la presente invención, por lo tanto, dar a conocer un procedimiento, en el que para una palabra a buscar predeterminada posibilita una búsqueda rápida y satisfactoria, de acuerdo con los datos que se encuentran en relación con la palabra a buscar, de manera que el procedimiento es tolerante con respecto a posibles  
40 desviaciones del concepto de búsqueda. Igualmente, es objetivo de la invención conseguir un sistema para la realización del procedimiento.

Estos objetivos se consiguen mediante el procedimiento de la reivindicación 1 y el sistema de la reivindicación 8.

45 La ventaja esencial del procedimiento objeto de la invención consiste en que la palabra a buscar es investigada según su significado, y los datos asociados al significado o al sentido se averiguan en la búsqueda. En este caso, se introducen en forma de listado en un archivo de buscadores los conceptos de búsqueda que tienen sentido como los llamados buscadores. Estos buscadores son morfemas de unidades de sentido mínimo, en especial, del idioma alemán.

50 Este procedimiento utiliza el hecho de que el número de estos morfemas se mantiene, en relación con las palabras que se pueden componer a base de ellos, en un ámbito susceptible de control.

55 Es ventajoso, aparte de los morfemas, introducir en el archivo de buscadores, como unidades adicionales, partes componentes de palabras definidas de modo propio (palabras base), para reducir la tasa de fallos en la búsqueda. Estas palabras base pueden estar compuestas, por ejemplo, de un morfema, tal como por ejemplo, "bord", y un añadido, tal como por ejemplo, "ell", de manera que el morfema, en este caso "bord", lleva a una clase de artículo no deseado, por ejemplo, a "Bordschuh", a pesar de que en este caso el usuario ha hecho la búsqueda de "Bordell". Para evitar esta confusión, es ventajosa la introducción de palabras base de este tipo.

60 De manera general, mediante el procedimiento según la invención, se consigue una unión o conexión física como, por ejemplo, por medio de Internet, entre un ordenador utilizado por el usuario y un servidor, en el que se pueden encontrar informaciones interesantes para el usuario, por ejemplo, en un archivo de información. De esta manera, se consigue una asociación de datos, no correlacionados hasta el momento, y se abre la posibilidad de facilitar  
65 informaciones.

- 5 Dado que los buscadores en forma de morfemas, palabras base y otras unidades tales como morfemas funcionales, interfieren entre sí dentro de la palabra a buscar (ver "bord" y "bordell"), es ventajoso, en caso de existencia de varios buscadores, facilitar para la misma sección de la cadena de caracteres lingüísticos (en este caso, "bord"), el archivo asociado al buscador más largo (en este caso, "bordell") y/o el indicador ("Pointer") del archivo asociado. En la palabra a buscar, se escogen, por lo tanto, las partes conjuntas más grandes para limitar la búsqueda y constituir la de manera más eficiente. Cuando, de acuerdo con el procedimiento, se ha encontrado un buscador de orden superior, el buscador de orden inferior más corto será abandonado.
- 10 En la búsqueda, la palabra a buscar será facilitada a una unidad de comparación que investiga dicha palabra con respecto a la existencia de buscadores listados en el archivo específico y que tienen sentido, de manera que cuando se encuentra un buscador es facilitado a una unidad de salida para proceso posterior, un archivo asociado a dicho buscador y/o una indicación (Pointer) de un archivo asociado al buscador.
- 15 Además de los llamados buscadores, de la mayoría de los cuales existen desviaciones con el mismo sentido construidas según reglas fijas, se toman en consideración de manera adicional y ventajosamente, los morfemas funcionales antes explicados, que están dispuestos antes y/o después de los morfemas, y que, en sí mismos, no tienen ningún sentido propio. El número de morfemas funcionales del idioma alemán es muy limitado.
- 20 El procedimiento de la invención filtra de la palabra a buscar el buscador, en especial, teniendo en cuenta los morfemas funcionales y también, en ciertos casos, sucesiones de letras no identificables y facilita un archivo o referencia. El procedimiento de la invención se puede realizar de manera especialmente ventajosa con un ordenador, de manera que se pueden utilizar las correspondientes posibilidades de entrada y salida para la entrada de las palabras a buscar y la salida de las informaciones buscadas. De manera ventajosa, el procedimiento puede ser utilizado para su funcionamiento en una "máquina de búsqueda" en Internet.
- 25 Es especialmente ventajoso, que la búsqueda según la invención en base al buscador, en especial, de las palabras base y morfemas, conduce con elevada probabilidad a un resultado rápido y dotado de sentido. De este modo, el procedimiento se caracteriza por una gran tolerancia a inexactitudes en la entrada y a combinaciones de palabras, especialmente del idioma alemán, reunidas a base de una serie de conceptos individuales. Si bien la búsqueda de una palabra a buscar de este tipo combinado carecía de resultado con los procedimientos conocidos hasta el momento porque no existía ninguna reproducción exacta de la palabra, la búsqueda, según la invención, a base del buscador, conduce a un resultado satisfactorio que ayuda a finalizar la búsqueda.
- 30 De manera ventajosa, la búsqueda, según los buscadores, en una palabra de búsqueda, no se interrumpirá ya después de encontrar un único buscador, sino que se buscan otros buscadores. De esta manera, se descubren en ciertas circunstancias en una palabra a buscar varias unidades con sentido, de manera que es posible una referencia a un correspondiente número de archivos. El contenido de los archivos encontrados puede ser manipulado cuantitativamente y se puede trabajar con los mismos mediante los conocidos enlaces de Boole. Al encontrar un morfema funcional en una forma de realización ventajosa del procedimiento, se puede recurrir al
- 35 archivo asociado al correspondiente morfema.
- 40 De manera ventajosa, se encuentra en el archivo de buscador una gran cantidad de buscadores, por ejemplo, reunidos en orden alfabético. Esta lista debe ser establecida solamente una vez, y se puede utilizar para diferentes aplicaciones. Según la necesidad, se pueden establecer varias listas, de manera que cada una de las listas esté asociada a una rama profesional determinada. En este caso, es ventajoso asociar a cada morfema, dentro de dicha lista, una dirección del tipo de un indicador o "Point", que se refiere a un archivo asociado al buscador. También los morfemas funcionales pueden ser reunidos en archivos constituidos en forma de una o varias listas.
- 45 Es especialmente ventajosa la utilización del procedimiento, según la invención, en búsquedas en bases de datos, en las que están reunidos artículos de mercancías o servicios (a continuación, se indicarán de manera simplificada como artículos) de varias entidades ofertantes, por ejemplo, varios comercios individuales de una ciudad. En la búsqueda, después de la introducción de su palabra a buscar, el usuario consigue los artículos correspondientes al contenido de sentido de la palabra a buscar con las informaciones necesarias, tales como precio, entidad ofertante, y similares. Para simplificar esta búsqueda y optimizar su resultado, es ventajoso constituir el archivo asociado
- 50 como categoría, de manera que esta categoría represente el contenido de sentido del morfema. Así, por ejemplo, el morfema "Jugend" (juventud) podría ser asociado a la categoría "Alter 10-30 Jahre" (edad 10-30 años). Además, puede ser ventajoso que un artículo o servicio constituya su propia categoría directamente o con su concepto general constitutivo del tipo o clase.
- 55 Para optimizar la búsqueda dentro de los artículos contenidos en la base de datos, es ventajoso asociar cada categoría a un grupo principal subordinado. De esta manera, por ejemplo, la categoría "Alter 10-30 Jahre" (edad 10-30 años) puede constituir el contenido de un grupo principal "Altersgruppe" (grupo de edad), de manera que este grupo de edad puede tener como contenido, además, otras categorías tales como "Alter 0-10 Jahre" (edad 0-10 años). La categoría "edad 0-10 años" podría ser asociada, además, a los morfemas "Kind", "Klein" y "Spiel" ("Niño", "Pequeño" y "Juego").
- 60
- 65

Los siguientes grupos principales se han demostrado como posibilidad especialmente eficiente de clasificación, porque permiten clasificar una parte importante de todas las categorías imaginables dentro del campo de los artículos y servicios y, por lo tanto, los morfemas correspondientes a uno de estos grupos principales. En estos grupos principales, se trata de grupos de edad de clientes potenciales, grupos objetivo para artículos o servicios, y grupos de productos de artículos. La separación según estos grupos principales es ventajosa puesto que constituye una sistemática para la formación de un archivo subordinado (archivo de servicio) en el que se pueden reunir todos los artículos y servicios.

En la constitución de este archivo se introducen los artículos, de manera que cada artículo queda asociado a una frase determinada como archivo de característica. Este archivo de características presenta un número de lugares de almacenamiento correspondiente al número de grupos principales, de manera que en uno o varios de estos lugares de almacenamiento se puede introducir una categoría asociada al correspondiente grupo principal. De este modo, la entidad ofertante de los artículos está obligada a categorizar los artículos de manera apropiada a la búsqueda, según los buscadores. El archivo potencial comprende, por lo tanto, artículos dispuestos en filas, una debajo de la otra, de manera que cada artículo está dispuesto en uno de una serie de espacios correspondientes al número de grupos principales, de manera que los lugares de dichos espacios están cubiertos por una categoría que corresponde a uno de los grupos principales.

En caso de existir uno de estos archivos potenciales, es sencillo, en una búsqueda organizada, de acuerdo con la invención, en base a los buscadores, reunir, en una etapa del proceso las categorías asociadas a los buscadores y buscar este archivo sumatorio establecido de esta manera y asociado a la palabra a buscar, con la estructura de un archivo de características, dentro del archivo potencial. Cuando se ha encontrado un archivo de características con el contenido del archivo sumatorio, se pueden facilitar al usuario en una última etapa del procedimiento, los artículos y/o servicios cuyos archivos de características se corresponden, por el contenido, en el archivo sumatorio.

Es especialmente ventajoso, cuando el usuario, en este caso la entidad ofertante, se puede servir de una máscara de entrada para el envío del archivo potencial, mediante la cual pueda enviar de manera simple y confortable, mediante el texto de entrada no normalizado del archivo potencial, el artículo solicitado como mercancía. Para ello facilita, con intermedio de la máscara, un artículo especial (por ejemplo, VW-Golft) con una descripción del artículo correspondiente (vehículo deportivo). La descripción del artículo será investigada, entonces, en cuanto a la existencia de los buscadores ("sport", "fahr") ("sport", "móvil-"), y, al usuario, las categorías asociadas a los buscadores (por ejemplo, "Tuningwerkstätten" y "Autovermietungen") ("Talleres de tuneado" y "Alquiler de automóviles"). El procedimiento de la invención se puede utilizar de manera especialmente ventajosa en el llamado "cross-selling" ("venta cruzada"). Este tipo de venta se ha originado en relación a los sistemas de comercio electrónico ("E-Commerce") y se caracteriza por la asociación de determinadas mercancías a otras mercancías para que el cliente, al buscar una mercancía, se le ofrezca una mercancía adecuada como ampliación. El procedimiento, según la invención, se puede utilizar especialmente bien en bases de datos que comprenden la oferta completa de servicios y mercancías de una ciudad. De esta manera, dado que existe solamente un número limitado de categorías que presentan sentido, se pueden enlazar entre sí determinadas categorías (por ejemplo, ALTER 0-30 JAHRE y SPORT) (EDAD 0-30 AÑOS y SPORT). En este caso, es especialmente ventajoso posibilitar al usuario una búsqueda asociativa, de manera que éste perciba la búsqueda en Internet como un paseo por la ciudad y, por ejemplo, en el negocio del deporte, que es especialmente adecuado para el grupo objetivo de juventud, encuentre también un anuncio sobre el tema "excursionismo".

El procedimiento, según la invención, será descrito a continuación en base a un diagrama mostrado en la figura.

La figura muestra un esquema de la búsqueda, tal como es realizable en un ordenador. Como primera fase, se facilita la palabra a buscar 1. Esta palabra a buscar será comparada en un archivo (archivo de buscadores o lista de morfemas) 2 en otra etapa del procedimiento en la que existe un listado de una serie de buscadores 3, de manera que el número puede llegar de unos 10.000 hasta 20.000 morfemas y algunos centenares de palabras base en el caso del idioma alemán. En el esquema mostrado, los buscadores 3 se han representado como símbolos. En la comparación de la palabra a buscar 1 y del archivo de buscadores 2, se descubrirán, en la palabra a buscar, los buscadores "Kreis" ("Círculo") y "Dreieck" ("triángulo"). En las partes componentes "Schlange" 4 (serpiente), se trata de morfemas funcionales unitivos que, según el análisis, son eliminados como lastre, tal como se ha representado por la papelera 5. La lista de morfemas funcionales no se ha mostrado en esta representación. Los buscadores listados 3 son asociados por una referencia (Pointer) 6 a otro archivo 7.

Como ejemplo, se podría pensar en la búsqueda, según la palabra "Turnschuhe" (zapatillas). Como morfemas, se encontrarían, entonces, las partes componentes "Turn" y "Schuh", mientras que "e" sería eliminada como morfema funcional sin sentido.

El otro archivo 7 está dividido en espacios 8, de manera que cada espacio 8 representa un grupo principal. Cada línea 9 del archivo está asociada a un buscador, de manera que el indicador 6 muestra uno de cada buscador 3 de un lugar de almacenamiento del archivo 7. El contenido de un lugar de almacenamiento representa una categoría "A" o "B".

De la búsqueda de zapatillas, el indicador de "Schuh" (zapato) indicaría la categoría "BEKLEIDUNG" (vestido), de manera que la categoría "BEKLEIDUNG" podría estar asociada al grupo principal "grupo de producto". El indicador de "Turn" mostraría la categoría "SPORT", de manera que "SPORT" se asociaría al grupo principal "Zielgruppe" (grupo objetivo).

5 En otra etapa adicional del proceso, las categorías "A" y "B" que se han encontrado se unificarían en un archivo sumatorio 10, que corresponde, en su formato, a las líneas de un archivo potencial 11, por lo menos parcialmente. El archivo potencial 11 contiene todos los productos "P" predeterminados por la entidad ofertante, en especial, mercancías y servicios, de manera que a cada producto "P" están asociadas informaciones en un archivo de  
10 características 12. De esta manera, un archivo de características 12 constituye una línea del archivo potencial 11. El formateado del archivo de características 12 tiene lugar según la división en grupos principales. El contenido de un archivo de características 12 consiste en categorías "A" y "B". En otra etapa, el archivo sumatorio 10 será comparado con el archivo potencial 11 y, si existe correspondencia del archivo sumatorio 10 y un archivo de características 12, se facilitará el producto "P". Al producto "P" está acoplado un indicador 13, que hace referencia a la frase de datos con la información completa del producto. Esta información completa será facilitada, entonces,  
15 al usuario.

En el ejemplo práctico, la entidad ofertante había previsto sus zapatillas con la indicación a los grupos objetivo "SPORT" y al grupo de productos "Bekleidung" "VESTIDO", y los había facilitado al archivo potencial. Una  
20 comparación de archivo sumatorio y archivo potencial condujo al resultado de "zapatillas". Este resultado podría ser facilitado al usuario con todas las informaciones necesarias.

## REIVINDICACIONES

1. Procedimiento para establecer una conexión entre una palabra de búsqueda que consiste en caracteres lingüísticos mínimos y un archivo existente en un ordenador, en el que en una primera etapa del procedimiento se especifica la palabra a buscar, con intermedio de una unidad de entrada, en particular, un teclado, **caracterizado porque**, la palabra a buscar es examinada en cuanto a su significado al facilitarla a una unidad de comparación, que examina la palabra a buscar (1) en cuanto a la existencia de términos significativos de búsqueda, llamados "buscadores" (3), relacionados en un archivo de buscadores, de manera que el archivo de buscadores contiene morfemas en el sentido de unidades significativas mínimas en un idioma y, en el que se determinan los datos asociados a la significación, de manera que al encontrar un buscador (3), un archivo asociado al buscador (3) y/o un llamado indicador de referencia (6) es facilitado a una unidad de salida de un archivo asociado al buscador (3) para su elaboración posterior.
2. Procedimiento, según la reivindicación 1, **caracterizado porque** el archivo de buscadores contiene, además de los morfemas, palabras base definidas.
3. Procedimiento, según la reivindicación 1 ó 2, **caracterizado porque** la palabra a buscar (1) es examinada en cuanto a la existencia de otros buscadores (3), de manera que para cada buscador (3) encontrado, el archivo (7) asociado con este buscador (3) es enviado directamente o por medio de un indicador (6), de manera que, en particular, en el caso en el que existen varios buscadores para la misma sección de la cadena de características, se envía el archivo asociado con el buscador y/o indicador más largo al archivo asociado.
4. Procedimiento, según una de las reivindicaciones anteriores, **caracterizado porque** la palabra a buscar (1) es examinada en cuanto a la existencia de morfemas funcionales unidos (4) sin contenido significativo directo, de manera que un morfema funcional encontrado (4) es conectado con el morfema asociado (3), y el archivo (7) asociado con este morfema (3) es enviado directamente o por medio de un indicador (6).
5. Procedimiento, según la reivindicación 4, **caracterizado porque** cada uno de los buscadores (3) del archivo de buscadores (2) tiene un indicador (6) que indica el archivo (7) asociado con el buscador (3) acoplado al mismo.
6. Procedimiento, según una de las reivindicaciones anteriores, **caracterizado porque** el archivo asociado con cada buscador representa una categoría (A, B), de manera que, en particular, se forma una categoría (A, B) a partir de una clase de mercancías o una clase de servicios.
7. Procedimiento, según una de las reivindicaciones anteriores, **caracterizado por** una máscara de entrada, con la que el usuario, por medio de un texto de entrada, puede facilitar la archivo potencial (11) mercancías y/o servicios, de manera que un artículo especial que comprende la descripción de un artículo correspondiente, es introducido con intermedio de la máscara, de manera que la descripción del artículo es examinada en cuanto a la existencia de buscadores, y en el que al usuario se le sugieren categorías que están asociadas con los buscadores.
8. Sistema dotado de medios para implementar el procedimiento, según una de las reivindicaciones anteriores, que comprende, como mínimo, dos ordenadores, en el que se puede especificar una palabra a buscar, que consiste en los caracteres lingüísticos mínimos por parte del usuario con intermedio de una unidad de entrada del primer ordenador, de manera que un archivo de informaciones que contienen datos de interés para el usuario existe en la memoria del segundo ordenador, y en el que los dos ordenadores están conectados mediante una red, en particular, mediante Internet, **caracterizado por** una unidad de comparación realizada en uno de los dos ordenadores o en un tercer ordenador, de manera que la unidad de comparación determina el significado de la palabra a buscar al examinar la misma en cuanto a la existencia de buscadores significativos (3) relacionados en un archivo de buscadores, de manera que el archivo de buscadores como buscador, contiene morfemas en el sentido de las unidades lingüísticas significativas mínimas, y en el que los datos asociados con el significado de la palabra a buscar se determina por el hecho de que, cuando un buscador (3) ha sido encontrado, la información del archivo de información asociado con el buscador (3) y/o una referencia (6) a la información asociada con el buscador (3) y almacenada en el archivo de información, es enviada al archivo de salida para el proceso adicional.
9. Procedimiento, según la reivindicación 8, **caracterizado porque** el archivo de informaciones comprende una serie de mercancías y/o servicios.
10. Máquina de búsqueda con medios para la búsqueda de la oferta de mercancías y servicios en Internet utilizando el procedimiento, según una de las reivindicaciones anteriores, en la que el usuario recibe acceso a un registro de datos.

