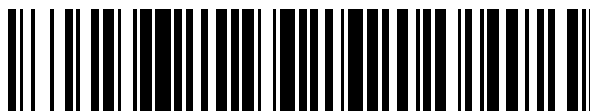


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 394 002**

51 Int. Cl.:

**G06F 17/30** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **05109402 .7**

96 Fecha de presentación: **10.10.2005**

97 Número de publicación de la solicitud: **1783633**

97 Fecha de publicación de la solicitud: **09.05.2007**

54 Título: **Motor de búsqueda para realizar una búsqueda referida a un lugar**

45 Fecha de publicación de la mención BOPI:

**04.01.2013**

45 Fecha de la publicación del folleto de la patente:

**04.01.2013**

73 Titular/es:

**SEARCHTEQ GMBH (100.0%)  
Wiesenhüttenstrasse 18  
60329 Frankfurt am Main, DE**

72 Inventor/es:

**HANTKE, REIMAR y  
LOHMEIER, FLORIAN**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

ES 2 394 002 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**

Motor de búsqueda para realizar una búsqueda referida a un lugar.

Campo de la invención

La presente invención concierne a un motor de búsqueda para realizar una búsqueda referida a un lugar.

## 5 Antecedentes de la invención

Los motores de búsqueda son ordenadores especiales o instalaciones informáticas programadas para la búsqueda por páginas web que satisfacen determinados criterios de búsqueda ingresados por un usuario. Para desarrollar su cometido, un motor de búsqueda carga las páginas de Internet en el ordenador del motor de búsqueda, indexa las páginas escudriñadas y proporciona también una interfaz de usuario y un mecanismo de consulta para filtrar las páginas indexadas respecto de criterios de búsqueda ingresados por el usuario e indicarle al usuario las páginas entonces encontradas, llamadas aciertos o éxitos (hits).

Para la descarga, un motor de búsqueda contiene típicamente una llamada araña (crawler) que contacta automáticamente con direcciones de Internet y descarga los contenidos de los sitios web correspondientes para su ulterior procesamiento (indexación).

15 No obstante, en contraste con los bancos de datos ya conocidos desde hace bastante tiempo, los contenidos de los sitios web son en general informaciones no estructuradas y el contenido del significado de los distintos conceptos (términos) de un sitio web sólo puede identificarse con dificultad. Esto restringe fuertemente las posibilidades de indexación y, por tanto, las posibilidades de búsqueda. Por tanto, una indexación de páginas web realizada por un motor de búsqueda es de hecho siempre una indexación de texto completo, es decir que a partir de todos los términos que se presentan en el sitio web (exceptuadas palabras de detención predefinidas y carentes de significado) se forma un índice de texto completo con el que se "empareja" después la petición de búsqueda.

20 En la tecnología convencional de los motores de búsqueda un usuario introduce en una interfaz de ingreso unos términos de búsqueda sobre la base de los cuales se envía después una petición de búsqueda (query) a un inventario de datos del motor de búsqueda, y una aplicación de los términos de búsqueda al inventario de datos o al índice da entonces como resultado eventualmente "coincidencias" o "aciertos" y se indican las páginas o vínculos correspondientes para el usuario.

Un problema de los motores de búsqueda convencionales consiste en que es difícil restringir los aciertos logrados a un criterio de búsqueda geográfico determinado. El usuario puede ciertamente ingresar como término de búsqueda un lugar, por ejemplo "Berlín", pero esto no significa que se encuentren solamente páginas que contengan la referencia geográfica deseada. Por el contrario, debido a la indexación de texto completo, que ciertamente no diferencia según contenidos de significados, se encuentran también páginas en las que la palabra "Berlín" no emerge como lugar de origen geográfico de la oferta de páginas web, sino con otro significado. Así, por ejemplo, con el ingreso de una búsqueda con los términos de búsqueda "taller de automóviles" y "Berlín" puede surgir también como acierto una página en la que alguien informe de su excursión a Berlín, en la que el coche sufrió daños durante la excursión y el interesado tuvo que buscar un taller de automóviles. Sin embargo, tales aciertos son poco deseables en el caso de una búsqueda referida a un lugar que contenga "Berlín" como criterio de origen geográfico.

30 La solicitud de patente internacional WO 01/63479 A1 se refiere a una codificación y representación espacial de información. En este contexto, la solicitud de patente describe un sistema informático que comprende un sistema de memoria que contiene informaciones en forma de documentos junto con informaciones espaciales sobre estos documentos. Este sistema informático comprende, además, dispositivos para la acumulación de datos, el análisis de datos, la búsqueda de datos, la representación de datos y los servicios de portal. El sistema informático comprende, además, una interfaz de mapa a través de la cual los usuarios pueden consultar el sistema de memoria y ver una representación de los resultados de la consulta presentados en un mapa.

45 La solicitud de patente norteamericana US 2004/0267723 A1 se ocupa de la reproducción de indicaciones con documentos con uno o varios temas mediante el empleo de informaciones de interés por temas. La solicitud de patente describe que pueden evitarse ambigüedades con respecto al interés principal de un usuario de modo que se pueda indicar una publicidad útil dirigida al interés principal del usuario. Tales ambigüedades pueden resolverse vigilando el comportamiento del usuario, determinando el interés principal del usuario en base al comportamiento vigilado y proporcionando la publicidad que esté dirigida al interés principal así determinado del usuario.

50 La solicitud de patente europea EP 1039265 A1 muestra un sistema de geoinformación que administra un banco de datos con lugares y proporciona una aplicación que accede a este banco de datos para presentarle a un usuario datos referidos a un lugar. Los usuarios pueden acceder a un servidor en el que está situado el banco de datos del sistema de información geográfica (banco de datos GIS). Este banco de datos proporciona también un sistema de realimentación mediante el cual un usuario puede confirmar la información del banco de datos GIS cuando éste haya

buscado un lugar en el mundo real.

Por tanto, un cometido de la presente invención consiste en proporcionar un motor de búsqueda que suministre como aciertos aquellas páginas de Internet para las cuales se cumple un criterio de origen geográfico deseado.

Sumario de la invención

- 5 Según un ejemplo de realización de la invención, ésta comprende un motor de búsqueda para realizar una búsqueda por páginas de Internet para los cuales se cumpla un criterio de origen geográfico ingresado por el usuario como término de búsqueda, presentando el motor de búsqueda:
- un dispositivo para escudriñar en una pluralidad de páginas de Internet;
- 10 un dispositivo para extraer datos geográficos de las páginas escudriñadas, designando los datos extraídos la respectiva referencia geográfica o la asignación geográfica de la página o del ofertante de la página;
- un dispositivo para formar un inventario de datos en el que están asociados a un gran número de páginas de Internet escudriñadas los respectivos datos geográficos extraídos de estas páginas de Internet;
- una interfaz de ingreso para ingresar una petición de búsqueda por el usuario, haciendo posible la interfaz de ingreso que el usuario ingrese un criterio de origen geográfico además de otros términos de búsqueda;
- 15 un escudriñamiento del inventario de datos y una entrega de las páginas de Internet para las cuales se cumplen el criterio de origen geográfico de los demás términos de búsqueda por cotejo con los contenidos de las páginas de Internet y los respectivos datos geográficos asignados a éstas.
- La extracción de informaciones de origen geográfico y su asignación a páginas individuales hacen posible la creación de un inventario de datos de páginas de Internet que pueden ser escudriñadas deliberadamente según
- 20 criterios de origen geográfico. El dispositivo para la extracción de información de origen geográfico según el ejemplo de realización antes citado presenta también:
- un dispositivo para aplicar un mecanismo de regulación al contenido de una página de Internet a fin de extraer las informaciones que, de conformidad con los criterios del mecanismo de regulación, pudieran especificar una dirección o un origen geográfico;
- 25 un dispositivo para verificar las posibles informaciones de origen geográfico por cotejo con un inventario de datos existente constituido por direcciones y/o partes de dirección existentes,
- en donde los datos de dirección especificados según los criterios del mecanismo de regulación para los que el cotejo con el inventario de datos existente de direcciones existentes no suministra ninguna coincidencia son retirados del inventario de datos que se debe escudriñar.
- 30 La aplicación de un mecanismo de regulación a los contenidos de páginas de Internet individuales hace posible la comprobación referente a si algunos componentes individuales satisfacen condiciones predeterminadas y, por tanto, entran en consideración como candidatos para informaciones de origen geográfico - por ejemplo, direcciones -, o bien suministra candidatos correspondientes. La comprobación o el cotejo de tales candidatos con un inventario de datos de dirección o partes de dirección existentes hace posible un incremento adicional de la probabilidad de que
- 35 los candidatos extraídos según el mecanismo de regulación sean datos de dirección. Eventualmente, se pueden desechar candidatos que han "fracasado" en esta prueba, con lo que quedan realmente tan sólo direcciones válidas o datos de origen geográfico válidos.
- Según un ejemplo de realización, el motor de búsqueda comprende también:
- un dispositivo para dictaminar si la página escudriñada es la página de un ofertante comercial.
- 40 La comprobación referente a si se trata de la página de un ofertante comercial hace posible que se recojan solamente páginas comerciales en el inventario de datos y, por tanto, se emitan éstas como resultados.
- Según un ejemplo de realización, el motor de búsqueda comprende también:
- un dispositivo para geocodificar la página de Internet escudriñada obteniendo una geocodificación por medio de un sistema de coordenadas geográficas en base a las informaciones de origen geográfico extraídas por cotejo con un
- 45 inventario de datos de geoinformación existentes para proporcionar una información de dirección extraída o una información geográfica extraída y asignando esta geocodificación a la página de Internet.
- La asignación de datos de geocodificación a las páginas web en base a las informaciones geográficas extraídas hace posible restringir la búsqueda exactamente a condiciones de búsqueda definidas por medio de coordenadas geográficas y emitir los aciertos correspondientes. Aparte de informaciones de lugar exactas, esto incluye

especialmente también la búsqueda del entorno o bien el empleo de una interfaz de ingreso en forma de un mapa sobre el cual se defina entonces el territorio de búsqueda.

Según un ejemplo de realización, el motor de búsqueda comprende también:

5 un dispositivo para escudriñar las distintas páginas de Internet según una pluralidad de términos que son adecuados para clasificar las páginas de Internet con arreglo al contenido ofrecido, asignándose a la página de Internet en el caso de un acierto - eventualmente aplicando otras condiciones - una clasificación correspondiente.

10 El escudriñamiento según términos de clasificación de un inventario predefinido de tales términos hace posible la asignación de tales términos de clasificación a las distintas páginas web. Esto puede emplearse después, por ejemplo, para producir una especificación adicional de la petición de búsqueda, por ejemplo para suministrar como aciertos solamente las páginas web a las que se ha asignado un término de clasificación ingresado por el usuario.

Según un ejemplo de realización, el motor de búsqueda comprende también:

un dispositivo para indexar las páginas escudriñadas a las que se han asignado informaciones de origen geográfico y eventualmente otras informaciones de clasificación;

un dispositivo para emparejar los términos de búsqueda con el contenido del índice formado;

15 una salida de los aciertos obtenidos, en donde las informaciones de origen geográfico asignadas a una página de Internet sirven, por cotejo con las informaciones de origen geográfico ingresadas como término de búsqueda, como criterio de filtro para la emisión de la lista de aciertos.

20 La indexación del inventario de datos de páginas web o direcciones de Internet a las que se han asignado informaciones de origen geográfico y eventualmente aún otras informaciones, tales como geocoordenadas y/o informaciones de clasificación, hace posible el cotejo con una petición de búsqueda que, entre otras cosas, contiene un criterio de búsqueda geográfico, así como la entrega de los aciertos correspondientes al usuario.

Según un ejemplo de realización, el motor de búsqueda comprende también:

25 un dispositivo para normalizar las informaciones geográficas extraídas a fin de poner éstas en un formato estandarizado que entregue después al usuario como "tarjeta de negocio" (business card) las informaciones de dirección y eventualmente otras informaciones de contacto junto con la dirección de Internet.

30 La normalización puede efectuarse sustituyendo informaciones que puedan presentarse en forma diferente, pero que tengan el mismo contenido de significado, por un término estándar o un formato estándar o bien transformándolas en éste, con lo que se le entrega siempre al usuario en el mismo formato, por ejemplo, una dirección o un número de teléfono. Mediante la entrega junto con el URL del acierto, el usuario puede reconocer enseguida informaciones especialmente importantes para él, tales como informaciones de dirección o un número de teléfono, y, en ciertas circunstancias, no es entonces absolutamente necesario que clique en el vínculo del acierto y vea la página correspondiente.

Según un ejemplo de realización, el motor de búsqueda comprende también:

35 un inventario de datos de direcciones de Internet que deberán ser descargadas por la araña y escudriñadas por el dispositivo de extracción,

un dispositivo para la adaptación dinámica del inventario de datos por:

agregación de nuevos vínculos que se han obtenido por el dispositivo de extracción al escudriñar las páginas descargadas; y/o

40 rechazo de direcciones de Internet en las que el dispositivo de extracción ha comprobado que no se cumplen los criterios prefijados para la extracción de una información geográfica; así como

un dispositivo para repetir la descarga y el escudriñamiento de las direcciones de Internet del inventario de datos.

45 La adaptación dinámica del inventario de datos hace posible recoger nuevos contenidos en el inventario de páginas web a investigar, así como retirar páginas no relevantes de éste, en caso de que estas páginas no correspondan a un criterio de relevancia predefinido (por ejemplo, considerar como relevantes solamente páginas de ofertantes comerciales).

Según un ejemplo de realización, un motor de búsqueda comprende:

un dispositivo para obtener informaciones adicionales a indicar que se indiquen adicionalmente a los aciertos

indicados para una petición de búsqueda, presentando este dispositivo:

un dispositivo para obtener el sector temático para el que deben indicarse informaciones adicionales.

5 La indicación de informaciones adicionales puede ser útil para el usuario cuando éstas estén temáticamente relacionadas con los aciertos. Éstas pueden ser útiles también para el operador del motor de búsqueda - por ejemplo, para la incorporación de publicidad temáticamente adecuada -.

Según un ejemplo de realización, un dispositivo para obtener el sector temático comprende:

un dispositivo para contar la frecuencia de algunas palabras individuales que se presentan en el acierto o aciertos, a fin de para obtener en base a las palabras que se presentan con la mayor frecuencia el sector temático para el que deberán indicarse informaciones adicionales; y/o

10 un dispositivo para consultar en sectores temáticos asociados a los respectivos aciertos a fin de obtener en base a ellos el sector o los sectores temáticos para los cuales se deben indicar informaciones adicionales.

El recuento de las palabras en los aciertos, en donde se realiza un recuento separado para cada palabra diferente, es un procedimiento eficiente para obtener el sector temático cuando se acceda a un léxico en el que las palabras estén asignadas a respectivos sectores temáticos.

15 Según un ejemplo de realización, las informaciones adicionales indicadas son vínculos publicitarios y la sucesión de indicaciones de vínculos publicitarios se basa en la frecuencia con la que ya se ha clicado en un vínculo publicitario. Esto hace posible una conexión y facturación eficientes de la publicidad para el operador del motor de búsqueda.

Breve descripción de los dibujos

20 La figura 1 muestra esquemáticamente una estructura de un motor de búsqueda según un primer ejemplo de realización de la invención.

La figura 2 muestra un diagrama de flujo que ilustra el funcionamiento de un motor de búsqueda según un ejemplo de realización de la invención.

La figura 3 muestra un diagrama de flujo que ilustra el funcionamiento de un motor de búsqueda según otro ejemplo de realización de la invención.

25 Descripción detallada

A continuación, se describe la presente invención en detalle con ayuda de varios ejemplos de realización.

La figura 1 muestra esquemáticamente una configuración según un primer ejemplo de realización de la presente invención.

30 En la figura 1 se representa esquemáticamente un motor de búsqueda según un primer ejemplo de realización. El motor de búsqueda se implementa por medio de un ordenador 100. El ordenador dispone de un enlace 110 con la Internet 120. Por medio de un dispositivo de escudriñamiento de Internet, una llamada araña 125, el ordenador 100 está en condiciones de descargar sistemáticamente diferentes páginas de Internet y escudriñar su contenido. La araña 125 almacena los contenidos en una memoria 130. Un dispositivo de extracción 135 escudriña las páginas descargadas para establecer si contienen informaciones de origen geográfico que aludan al origen geográfico del ofertante de las páginas o de la página web escudriñada. Si ocurre esto, se extraen entonces las informaciones geográficas correspondientes y se las asigna después a la página correspondiente. Las páginas para las cuales tuvo éxito esta asignación se almacenan después en un inventario de datos 140.

35 Una interfaz de ingreso 145 hace posible que el visitante ingrese como términos de búsqueda uno o varios términos de búsqueda 165 y, por separado, un criterio de origen geográfico 170. Los términos de búsqueda pueden ser, por ejemplo, términos temáticos (por ejemplo, "zapatos", "pizza", "restaurante", etc.). El criterio de origen geográfico puede ser, por ejemplo, un nombre de lugar, una parte de ciudad, un nombre de calle, etc. o bien - en caso de que la interfaz de ingreso y el motor de búsqueda estén asistidos como se describe aún más adelante - coordenadas de lugar o un lugar o entorno fijado por medio de un mapa. Esta petición de búsqueda es procesada después por el dispositivo de búsqueda 150 buscando para ello en el inventario de datos 140 las páginas de Internet a las que se aplican los términos de búsqueda. Las páginas de Internet correspondientes son emitidas después por medio de la interfaz de salida 150 e indicadas al usuario. El dispositivo 150 para escudriñar el inventario de datos emplea aquí las informaciones de origen geográfico que se han ingresado junto al uno o varios términos de búsqueda arbitrarios como criterio de búsqueda geográfico adicional 170, para filtrar y separar del inventario de datos 140 las páginas web a las que se aplica el criterio de origen geográfico. Según un ejemplo de realización, esto puede realizarse aplicando los términos de búsqueda arbitrarios 165 al inventario de datos por medio de una tecnología de motor de búsqueda convencional para localizar un primer juego de aciertos correspondientes, y de este primer juego de

aciertos se filtran y separan después, comprobando las informaciones geográficas asignadas a los aciertos, únicamente las páginas web y se emite un segundo juego de aciertos 175 que representan en último término el resultado de búsqueda y a los cuales se aplica el criterio de origen geográfico.

5 A continuación, se describe la invención en detalle con algo más de precisión conforme a otro ejemplo de realización. La figura 2 muestra un diagrama de flujo que ilustra el funcionamiento de un motor de búsqueda según un ejemplo de realización de la invención.

10 En un paso 200 la "araña" escudriña varias páginas de Internet. Esto puede realizarse, por ejemplo, proporcionando a la araña un número predefinido de páginas de Internet, por ejemplo en forma de una base de datos 205 que es descargada después por la araña y almacenada en una memoria 210 de dicha araña. Las páginas almacenadas en la memoria de la araña son analizadas después por medio de un proceso de extracción 215, y especialmente las informaciones de origen geográfico, por ejemplo en forma de direcciones, son extraídas de los datos almacenados en la memoria de la araña y asignadas a las respectivas páginas. Según el presente ejemplo de realización, se emplea aquí un procedimiento multietapa para la extracción de los datos.

15 Cotejando con una base de datos de dirección 220 y aplicando un mecanismo de regulación 225 se extraen datos de dirección de los contenidos de páginas almacenados en la memoria 210 de la araña. Esto puede efectuarse, por ejemplo haciendo que un inventario de datos predefinido actuante como base de datos de dirección 220, que contiene direcciones (por ejemplo, los datos de una guía telefónica, un directorio de ramos comerciales u otro listado de direcciones), sea comparado con los datos de la página web y, en caso de coincidencias, se extraigan los datos de dirección correspondientes. A este fin, se pueden buscar no sólo coincidencias de uno a uno, sino que, por el  
20 contrario, se puede perseguir también un enfoque heurístico empleando un mecanismo de regulación. Así, por ejemplo, según el mecanismo de regulación, cuando se encuentra la palabra "dirección" en el índice de una página y también se encuentra en las filas siguientes una cifra que está presente en un listado de números de distrito postal, se puede partir de la consideración de que se trata aquí de una dirección. De manera semejante, se puede buscar, por ejemplo, por la abreviatura "Tel." o "No. Tel.", pudiendo partirse entonces, en caso de localización de tal  
25 abreviatura, de que las cifras que siguen después incluyen un número telefónico. Para una comprobación adicional se puede cotejar también, según el mecanismo de regulación, una parte del número telefónico localizado con, por ejemplo, un listado de números de prefijos telefónicos. Puede tener lugar una valoración de las direcciones de tal manera que se identifique a partir de varias direcciones contenidas en la página web aquella dirección que caracteriza la sede principal de una empresa o bien determina sus filiales o delegaciones exteriores. Esto se efectúa  
30 evaluando el entorno semántico con el lugar hallado de la dirección, así como comparando la frecuencia del lugar de aparición de tipos de dirección determinados. Además, se pueden identificar así también direcciones que no pertenezcan a la propia firma, sino que pertenezcan, por ejemplo, a proveedores de servicios o clientes.

35 De manera semejante, por medio del mecanismo de regulación se puede buscar por nombres de lugares y ciudades, con lo que, cuando preceda a estos un número de código postal, lo que puede comprobarse por cotejo con la base de datos 220, se deberá tratar con alguna seguridad de una parte de una información de dirección. El empleo del banco de datos 220 en combinación con un mecanismo de regulación correspondiente 225 hace posible entonces extraer informaciones de dirección de los datos de páginas web almacenados en la memoria de la araña.

40 El mecanismo de regulación puede contener también una comprobación referente a si la página escudriñada es una página de un "ofertante comercial" o bien una página no comercial. Puntos de arranque para la existencia de una página comercial serían, por ejemplo, la indicación de un enlace bancario, la mención de la forma jurídica de una firma (por ejemplo, GmbH, AG, GmbH & Co., etc.). Empleando tales mecanismos de regulación se puede adoptar entonces una decisión referente a si la página escudriñada es una página comercial o una página no comercial. En caso de que deban recogerse solamente páginas comerciales en el inventario de datos, la presencia de una página no comercial puede conducir entonces al rechazo de esta página.

45 Además de la extracción de datos de dirección, el mecanismo de regulación puede estar adaptado también de modo que se extraigan otros datos relevantes de una página web, por ejemplo el horario de apertura. Esto puede efectuarse también buscando por términos predefinidos (por ejemplo, "abierto" u "horario de apertura" u "horario comercial") y sometiendo luego los términos siguientes a un comprobación de plausibilidad o de formato para establecer si se trata de horarios de apertura. Esto puede efectuarse por cotejo con modelos predefinidos (plantillas)  
50 que sean representativos de posibles representaciones de horarios de apertura. Así, por ejemplo, se pueden depositar modelos para los días de la semana (Lu Ma Mi Ju, etc.) o modelos para las horas del reloj (un número de dos cifras, luego dos puntos o un punto y luego nuevamente un número de dos cifras). Por cotejo con estos modelos o "plantillas" se puede establecer si los datos extraídos son horarios de apertura.

55 En otro paso se "geocodifican" después las páginas de la memoria de la araña. Con geocodificación se quiere dar a entender aquí la asignación de datos geográficos a la página web, por ejemplo datos en forma de un grado de longitud y un grado de latitud u otros sistema de coordenadas comparable (X, Y). A este fin, se accede a las informaciones de lugar extraídas, que se presentan ciertamente ya como resultado de la extracción de dirección, por medio de un mecanismo de regulación 235 que se sirve a su vez de una base de datos 230 en la que están depositadas las coordenadas geográficas de los lugares. Así, por ejemplo, para una dirección de Munich en la

Maximilianstrasse 28 el proceso de geocodificación 235 puede acceder entonces a la base de datos 230 en la que están depositadas las coordenadas correspondientes X,Y para este lugar, por ejemplo como información en grados de longitud y de latitud. Estas coordenadas se asignan después a la página web correspondiente.

5 Según otro ejemplo de realización, el proceso de extracción puede incluir también la extracción de datos de clasificación. Con datos de clasificación se quiere dar a entender aquí datos que clasifican una página web de la memoria de la araña atendiendo a un contenido semántico u otro determinado. Una posibilidad de esta clasificación sería, por ejemplo, una clasificación de negocios que realice una clasificación en ramos diferentes o bien que realice una asignación a productos o marcas. Términos de clasificación adecuados pueden ser, por ejemplo, designaciones de ramos (moda, foto, publicidad, gastronomía, etc.) u otras categorizaciones (por ejemplo, en un hotel el número de  
10 estrellas).

Para realizar esta extracción de clasificación puede estar prevista una base de datos 240 que escudriñe en combinación con un mecanismo de regulación 245 las páginas almacenadas en la memoria de la araña para determinar si éstas pertenecen a un categoría de negocia determinada. Así, por ejemplo, la base de datos puede contener el término "taller de automóviles", y cuando este término surja después en una página escudriñada, se  
15 puede asignar entonces a esta página la clasificación "taller de automóviles".

Para evitar asignaciones erróneas, el mecanismo de regulación 245 puede estar construido aquí en forma redundante o compleja de modo que se puedan satisfacer varios criterios para asignar una clasificación de negocio determinada a una página web. Así, por ejemplo, se pueden escudriñar por separado partes diferentes de una página web (por ejemplo, título, cuerpo, encabezamiento, descripción, metaetiquetas, etc.), y solamente en el caso  
20 de la presencia de un acierto en partes diferentes de la página web se decide positivamente que se deberá conferir la clasificación correspondiente. Asimismo, puede estar prefijado también, por ejemplo, un valor umbral determinado que indique la frecuencia en que tiene que surgir al menos el término buscado para asignar una clasificación correspondiente a esta página web. Este valor umbral puede ser definido por separado para partes individuales del sitio web y adicionalmente para el sitio web en su totalidad, y únicamente al sobrepasarse todos los límites de valor  
25 umbral se asigna entonces la clasificación correspondiente. En términos muy generales, puede decirse que se escudriñan por separado partes diferentes del sitio web, se ponderan por separado los respectivos aciertos (también eventualmente respecto del número) y, finalmente, se combinan estos para dar una puntuación total que sirve entonces como fundamento de decisión (por ejemplo, por comprobación de si la puntuación total es mayor que un valor umbral determinado) referente a si se llega a conferir a la página la clasificación correspondiente.

30 Según un ejemplo de realización, se pueden asignar entonces varias clasificaciones de negocio a una página web. La base de datos 240 puede formarse entonces, por ejemplo, con ayuda de bancos de datos predefinidos, por ejemplo un directorio de ramos o bien una compilación de datos de un ofertante comercial que contenga diferentes términos de clasificación potenciales, para lograr una cobertura correspondientemente amplia con posibles clasificaciones de negocio.

35 Eventualmente, la clasificación puede estar subdividida también en planos jerárquicamente diferentes, con lo que puede tratarse de una taxonomía compleja cuyos elementos individuales se asignan cada uno de ellos en el caso de un dictamen positivo (es decir, se presenta la clasificación correspondiente para el mecanismo de regulación) a la página web correspondiente.

40 Como consecuencia de la aplicación del proceso de extracción 215 a las páginas web almacenadas en la memoria 210 de la araña se obtiene un inventario de datos 250 de páginas web a las que están asignadas varias informaciones adicionales correspondientes. Un ejemplo de un extracto del inventario de datos 250 se representa seguidamente en la tabla 1.

Tabla 1

URL	Dirección	Teléfono	Geodatos	Clasificación
<a href="http://www.automeier.de">www.automeier.de</a>	Hansestrasse 5 80331 Munich	089-2345	X, Y	Taller de automóviles

45 En la tabla 1 se representa tan sólo esquemáticamente un posible ejemplo para un juego de datos tal como éste podría resultar del proceso de extracción según la presente descripción. Cabe consignar a este respecto que se han listado únicamente a título de ejemplo la dirección y el número telefónico en columnas separadas para hacer posible aquí eventualmente también una indexación y búsqueda separadas. Sin embargo, los números telefónicos podrían estar contenidos también, por ejemplo, en los propios datos de dirección.

50 Además, cabría hacer notar todavía que únicamente está representado una clasificación de negocio en la tabla 1, pero que podrían estar previstas también, por ejemplo, varias clasificaciones de negocio, por ejemplo también

clasificaciones de negocio de planos de jerarquía diferentes, tal como, por ejemplo, a) automóvil, b) alquiler de automóviles, c) comercio de automóviles, d) reparación de automóviles, que llenarían entonces casillas en columnas diferentes de la tabla 1 para obtener una página web correspondiente.

5 El resultado del proceso de extracción 215 es entonces el inventario de datos 250 de páginas web con indicaciones de origen geográfico asociadas (por ejemplo, en forma de direcciones), tal como se representa a título de ejemplo en las dos primeras columnas de la tabla 1, eventualmente todavía con geodatos también asignados y con una o varias clasificaciones de negocio (según una taxonomía prefijada).

10 El inventario de datos 205, que es descargado por la araña y analizado por el proceso de extracción, se puede variar dinámicamente. A este fin, por ejemplo en el proceso de extracción 215, aparte de la búsqueda por datos de origen geográfico, se puede realizar también una búsqueda por vínculos en la página web, un proceso que no está representado gráficamente en el mecanismo de extracción 215 de la figura 2. Si se encuentra un vínculo de esta clase que remite a otra página web, se puede añadir entonces este vínculo al inventario 205 de direcciones a escudriñar, con lo que se varía este inventario dinámicamente por recogida de tales vínculos localizados. De manera semejante, se retiran del inventario de datos 205 a escudriñar direcciones de dicho inventario 205 para las que no se pudo verificar ningún origen geográfico o para las que no se cumplen otros criterios que son necesarios para la recogida en el inventario de datos (por ejemplo, la presencia de una página comercial). Esta variación dinámica del inventario de datos 205 hace posible que el inventario de datos resultante 250 se adapte dinámicamente a variaciones en Internet. Hay que hacer notar también a este respecto que, naturalmente, deberá realizarse repetidamente con este fin el proceso completo de escudriñamiento y extracción.

20 En base al inventario de datos 250 resultante y, dado el caso, dinámicamente variable se puede realizar entonces una búsqueda basada en una petición de búsqueda de un usuario. Esto se describe seguidamente en detalle con algo más de precisión.

A continuación, se describe el procesamiento de una petición de búsqueda haciendo referencia a la figura 3. Partiendo del inventario de datos 350, que es el resultado del arrastre de la araña y el proceso de extracción que se ha descrito con referencia a la figura 2, se indexa este inventario de datos. A este fin, las informaciones presentes en la página (por ejemplo, metaetiquetas, títulos, rótulos, texto puro, la relación de vínculos a texto), es decir, el número de los vínculos en un URL, en comparación con las palabras no vinculadas, así como también las informaciones que los describen (por ejemplo, designador dentro de un URL; número, designación y fuentes de los vínculos que remiten a esta página) son extraídas y pueden ser almacenadas en uno o varios índices separados. Estos uno o varios índices separados forman entonces el índice local 355 representado en la figura 3. El índice 355 puede estar formado de la manera convencional sobre la base de las páginas web del inventario de datos. Esto significa que, por ejemplo, se forma un índice de texto completo sobre las páginas del inventario 355, estando asignado el sitio web correspondiente a cada término del índice así formado. Sin embargo, además de la dirección del sitio web, está asignada también a un termino que está contenido en el índice la información geográfica que pertenece a este sitio web y se ha extraído en el proceso de extracción. Un fragmento de un índice así formado puede tener el aspecto que se representa seguidamente en la tabla 2.

Tabla 2: Fragmento del índice local

Término del índice	URL	Dirección
Habitación	<a href="http://www.hotelmaier.de">www.hotelmaier.de</a>	Karlplatz 5 80333 Munich
Habitación	<a href="http://www.hotelzimmer.de">www.hotelzimmer.de</a>	Sanderstrasse 3 90211 Würzburg
Zimmes	<a href="http://www.autohauszimmes.de">www.autohauszimmes.de</a>	Rothstrasse 3 80231 Munich

40 Aparte de un índice de texto completo, que indexa el contenido total del sitio web, se pueden proporcionar en el índice local, como ya se ha mencionado, otros varios índices (parciales) separados que se formen únicamente a base de partes determinadas del sitio web. Así, se pueden proporcionar índices separados para metaetiquetas, para títulos, para rótulos, para texto puro, para la relación de vínculos a texto, es decir, el número de vínculos en un URL, en comparación con las palabras no vinculadas, pero también para las informaciones que los describen (por ejemplo, designador dentro de un URL; número, designación y fuentes de los vínculos que remiten a esta página).

45 Cada uno de estos índices (parciales) representa entonces en cierto modo un fragmento determinado de los distintos sitios web. Los diferentes índices (parciales) pueden emparejarse entonces individualmente con la petición de búsqueda, las distintas peticiones (parciales) suministran resultados (parciales) y a partir de estos resultados parciales que incluyen en cada caso como aciertos ninguno, uno o varios sitios web, se puede formar después un resultado total, tal como se describe más adelante con mayor precisión.

50 Por tanto, el índice local 355 puede consistir en uno o varios índices (parciales). A continuación, se describe un



ejemplo de realización en el que el índice local consiste en solamente un índice (parcial), por ejemplo un índice de texto completo.

Si se aplica una petición de búsqueda al índice local 355 por medio de términos de búsqueda convencionales 360, esto puede realizarse entonces por medio de una tecnología de motor de búsqueda clásica que se ha representado a título de ejemplo como dispositivo de petición 365 en la figura 3. La aplicación de los términos de búsqueda 360 al índice suministra entonces un primer juego de resultados (sitios web o URLs) 370 que se filtran después nuevamente por medio de un dispositivo de filtrado posterior 375, concretamente aplicando el criterio de búsqueda más amplio de las informaciones de origen geográfico 380 que se han ingresado por el usuario por separado y adicionalmente a los términos de búsqueda normales 360. Únicamente los sitios web de la lista de resultados 370 para los cuales el dispositivo de filtrado posterior 375, comprobando la estructura de datos según la tabla 2, establece que se ha cumplido el criterio de origen geográfico según los parámetros de petición de búsqueda 380, forman entonces la lista de resultados definitiva 385 que se entrega al usuario. Así, por ejemplo, al ingresar el término de búsqueda "habitación" sin criterio geográfico adicional, se encuentran como primer grupo de aciertos las dos primeras líneas de la tabla 2. Al ingresar "Würzburg" como criterio de origen geográfico se filtra y separa entonces de los dos aciertos, como resultado definitivo, el término "habitación de hotel", ya que solamente este sitio web satisface el criterio de origen geográfico.

Según un ejemplo de realización, la entrega puede realizarse por medio de una interfaz de salida 390 que le indica al usuario los resultados en forma preferiblemente ordenada según su relevancia, por ejemplo en una secuencia que es el resultado de un "llamado proceso de ordenación por rangos". Una posibilidad para la realización de un proceso de ordenación por rangos es, por ejemplo, el empleo del llamado procedimiento de rango de página que se describe en la patente US 6.285.999. Si no se realiza esta ordenación por rangos, se le pueden presentar también al usuario los resultados 385 en forma desordenada. Además, es posible una entrega de todos los resultados atendiendo puramente a criterios geográficos, por ejemplo todas las firmas (de una determinada categoría) que residen en una calle.

Según un ejemplo de realización ventajoso, se indican los vínculos con los sitios web identificados como aciertos junto con la información de dirección extraída que se ha asignado a este vínculo como resultado del proceso de extracción 250. Eventualmente, se pueden indicar también por medio de la interfaz de salida 390 otras informaciones adicionales que resulten del proceso de extracción, por ejemplo la clasificación de negocio y/o también los datos geográficos en forma de coordenadas.

A continuación, se describe un ejemplo de realización en el que el índice local consiste en varios índices (parciales). Por tanto, según este ejemplo de realización se forman para el inventario de datos 350 varios índices (parciales), concretamente índices sobre categorías diferentes, pudiendo ser las categorías, por ejemplo, las informaciones adicionales que resultan del proceso de extracción, o bien índices que se forman a base de partes diferentes de los sitios web (como se ha descrito antes, por ejemplo, metaetiquetas, descripción, etc.). Estos varios índices pueden emplearse después para realizar una ordenación por rangos de los aciertos, tal como se describe seguidamente con más precisión. Las categorías que pueden agregarse para la ordenación por rangos pueden ser aquí, por ejemplo, los diferentes apartados de un sitio web, por ejemplo el título (title), el cuerpo (body), la descripción del sitio web (description), la cabeza (head), la información de vínculos, etc. Así, se pueden formar diferentes índices parciales para las distintas partes de la página web, y estos pueden emparejarse después con aciertos mediante la petición de búsqueda. Resultan entonces para los diferentes índices parciales diferentes listas de aciertos parciales que se combinan después una con otra, por ejemplo por formación de la cantidad media de todos los aciertos.

Los aciertos en los diferentes índices parciales pueden emplearse después también para calcular una "puntuación" representativa de la relevancia, por ejemplo estableciendo para los aciertos en los diferentes índices parciales una respectiva puntuación individual que se base en el número de aciertos en un índice determinado (es decir, la frecuencia con la que se presenta el término de búsqueda en el sitio web encontrado como acierto o en la parte del mismo que se ha empleado para la formación del índice), ponderándose entonces de manera diferente estas puntuaciones individuales según un esquema predefinido y añadiéndose luego las puntuaciones parciales de esta página web que se han obtenido para ella en las diferentes confrontaciones con los diferentes índices parciales. Se pueden realizar así conforme a los respectivos aciertos para los diferentes índices diferentes ponderaciones para las distintas categorías (body, title, description, link, etc.) que arrojan entonces por medio de un algoritmo para la combinación de los diferentes resultados un resultado total para una respectiva página que representa la relevancia de esta página. Así, por ejemplo, un acierto en el título de la página puede ponderarse con más fuerza que un acierto en el cuerpo (body), mientras que se puede adscribir una ponderación media a un acierto en la parte description. Los factores detallados pueden adaptarse y variarse de conformidad con las circunstancias.

Para dar un ejemplo concreto, el índice local puede consistir, por ejemplo, en un índice de texto completo y en un índice que se ha formado solamente por medio de la parte "Title" de sitios web. El sitio web [www.mamas-pizza.de](http://www.mamas-pizza.de) tiene entonces, por ejemplo, en el título la palabra "Mama", concretamente tan sólo una vez, pero doce veces en el texto completo total.

Sea el término de búsqueda "Mama" que se confronta con el índice parcial "Título" y con el índice parcial "Texto

completo". En ambos casos resulta como un acierto (aparte de presumiblemente otros muchos sitios web) el sitio web [www.mamas-pizza.de](http://www.mamas-pizza.de), es decir que éste surge como acierto en la lista de aciertos del índice de texto completo y también del índice del título. Se establece ahora para el acierto [www.mamas-pizza.de](http://www.mamas-pizza.de) en el índice del título una puntuación que, por ejemplo, se basa en la frecuencia en que se presenta el término de búsqueda (es decir, "Mama") en el título del sitio web [www.mamas-pizza.de](http://www.mamas-pizza.de). Dado que esto ocurre una sola vez, la puntuación de [www.mamas-pizza.de](http://www.mamas-pizza.de) para el índice parcial "Título" es igual a 1. Para el índice parcial "Texto completo", en el que ya ha aparecido ciertamente también [www.mamas-pizza.de](http://www.mamas-pizza.de) como acierto, se establece ahora también una puntuación. Sin embargo, dado que en el texto completo aparece 12 veces el término de búsqueda, esta puntuación es ahora de 12. Se ponderan ahora las dos puntuaciones, suponiéndose que la ponderación para "puntuaciones del título" es 3 y la ponderación para puntuaciones del texto completo es 1. Resulta así como puntuación total  $3 \times 1 + 1 \times 12 = 15$  para [www.mamas-pizza.de](http://www.mamas-pizza.de).

Se establece ahora también una puntuación correspondiente para toda las demás páginas web que se han encontrado como aciertos en esta petición de búsqueda. En base a las puntuaciones totales así obtenidas se pueden entregar después los aciertos ordenados según su relevancia establecida por la puntuación total.

Por tanto, se puede formar de la manera así descrita el índice local 355 a base de varios índices parciales que se combinan después uno con otro para formar en cierto modo un "índice total local".

En otro ejemplo de realización se pueden formar índices separados para una búsqueda de contenidos (una búsqueda según el "qué" y una búsqueda local (una búsqueda según el "dónde"). A este fin, se indexan los datos georreferenciados, concretamente de modo que se forme el índice sobre los datos locales, es decir que mediante el ingreso de los datos locales (por ejemplo, mediante las coordenadas correspondientes) se puede buscar por sitios web correspondientes a estos datos locales. Los datos locales pueden ingresarse aquí como coordenadas o bien en otra forma (por ejemplo, como nombres de lugares o direcciones, como marcaciones en un mapa, etc.). Cuando se ingresan en forma de direcciones o por medio de un mapa de ingreso, por ejemplo clicando el lugar de búsqueda o definiendo un entorno para la búsqueda del entorno, se transforma entonces este ingreso en coordenadas correspondientes que forman la base para la indexación y que pueden confrontarse después como petición de búsqueda con el índice.

Se puede confrontar entonces una petición de búsqueda "dónde" con el índice de coordenadas de lugar y se pueden filtrar posteriormente los resultados correspondientes empleando el otro criterio de búsqueda "qué", es decir que se filtran y separan de los aciertos para los que la petición de búsqueda de lugar se tradujo en un éxito, aquellos a los que se aplica entonces también el criterio de búsqueda "qué". A este fin, aparte de la clave primaria consistente en las coordenadas de lugar y las direcciones web asignadas, el índice contiene también las informaciones "qué" correspondientes, es decir, por ejemplo, el texto completo del sitio web correspondiente. En base a esto, se puede comprobar después si se aplican tanto el criterio "dónde" como el criterio "qué".

Un modo de proceder más eficiente al presentarse tanto el criterio de búsqueda "qué" como el criterio de búsqueda "dónde" consiste en realizar primero la búsqueda "qué" por medio del índice local y filtrar después nuevamente los aciertos encontrados mediante una comprobación referente a si se aplica para ellos el criterio de búsqueda "dónde". Por este motivo, esto es más eficiente que una búsqueda primero según el dónde y luego un filtrado posterior según el "qué", ya que cada sitio web tiene asignada solamente una coordenada de lugar que tiene que comprobarse en el filtrado posterior, pero cada sitio web contiene típicamente muchísimas informaciones "qué" (texto, etc.), lo que dificulta netamente un filtrado posterior según el criterio de búsqueda "qué" frente a un filtrado posterior según el "dónde".

Además, es posible una combinación de varios índices de modo que, por ejemplo, se puedan mezclar índices georreferenciados de, por ejemplo, diferentes estados federales con índice no georreferenciados como un índice de motor de búsqueda estándar. Se pueden mezclar así, por ejemplo, informaciones de lugar de léxicos no georreferenciables con las informaciones de un índice georreferenciado a una ciudad de conformidad con el valor de rango. Esto significa entonces que el índice local consiste, por ejemplo, en entradas que están georreferenciadas y en entradas que no están georreferenciadas. Una petición de búsqueda se traduce entonces, por ejemplo, tanto en aciertos georreferenciados como aciertos no georreferenciados que se valoran ciertamente después de manera diferente en la ordenación por rangos, pero que sin embargo, se entregan en una lista de resultados común.

El usuario puede definir una petición de búsqueda ingresando una palabra de búsqueda ("QUÉ"), pero como alternativa o adicionalmente también ingresando un territorio de búsqueda ("DÓNDE"; por ejemplo, lugar, región, calle). Este territorio de búsqueda puede establecerse, por ejemplo, por medio de un nombre, pero también por medio de un fragmento de mapa libremente seleccionable.

La lógica de búsqueda suministra entonces unos resultados correspondientes para el usuario mediante una comparación de ambos términos con las informaciones indexadas. Como ya se ha mencionado, esto ocurre en el caso de que se hayan definido tanto un criterio "qué" como un criterio "dónde", preferiblemente efectuándose primero un emparejamiento con el índice "qué" local que se ha formado por medio de los contenidos de página, y realizando a continuación un filtrado posterior de los aciertos respecto del criterio de búsqueda "dónde". Resulta

entonces un número de aciertos que se entrega al usuario.

Las condiciones de búsqueda que tienen que cumplirse para la entrega de un acierto pueden definirse, por ejemplo, de la manera siguiente:

- 5
- cuando se busca en un lugar, calle, región, se entregan solamente resultados que cumplen suficientemente este criterio
  - se puede buscar en el entorno de una posición (lugar, centro del lugar, dirección,...) a una distancia determinada (por ejemplo, radio).

Según un ejemplo de realización, se asiste al usuario en la búsqueda por medio de procedimientos de asistencia que le llevan con la rapidez de la ayuda a su resultado deseado. Esto puede comprender, por ejemplo,

- 10
- Posiciones auxiliares en caso de errores de teclado generando propuestas que son semejantes en modo de escritura a la palabra de búsqueda o lugar de búsqueda ingresado,
  - Posiciones auxiliares de tal manera que se le indiquen ya durante el ingreso del lugar los lugares o calles que coincidan con las letras ya ingresadas,
  - Una posición auxiliar tal que se le ofrezcan temáticamente búsquedas similares indicando para ello también, por ejemplo, términos de búsqueda similares sobre la base de un léxico,
  - Además, se le pueden proporcionar al usuario indicaciones sobre temas que se presentan con especial frecuencia en los resultados. A este fin, se establecen, por ejemplo, las palabras más frecuentes que se presentan en las páginas de acierto, se eliminan "palabras de detención" sin significado (por ejemplo, el, la, lo, un, etc.) y luego se comparan las palabras más frecuentes con un catálogo predefinido de temas a los que se han asignado respectivas indicaciones correspondientes. Si existe una coincidencia entre, por ejemplo, la palabra más frecuente y una entrada de esta clase en el catálogo (por ejemplo, la palabra hotel), se entregan también automáticamente las informaciones asignadas a esta entrada. Esto puede efectuarse entonces, por ejemplo, en la forma de "encontrará usted más informaciones sobre el tema hotel en...". Este mecanismo puede emplearse también para la incorporación de publicidad. En este caso, se han asignado entonces respectivos anuncios publicitarios predefinidos (típicamente varios de ellos), por ejemplo, a las entradas del catálogo, es decir, por ejemplo, a términos tales como taller de automóviles, panadería, librería, etc. Se indican entonces para un acierto los anuncios publicitarios (o un número determinado de ellos, por ejemplo, 3, 5 u 8).
- 15
- 20
- 25
- 30
- 35
- Los anuncios publicitarios asignados a un término del catálogo que son intercalados y pagados, por ejemplo, por clientes publicitarios del operador del motor de búsqueda, pueden ser sometidos entonces a una ordenación por rangos. Existen aquí distintas posibilidades para diferentes ejemplos de realización. Así, por ejemplo, la ordenación por rangos se puede basar en la frecuencia con que un anuncio es clicado por el usuario. Así, cada anuncio puede llevar asignada una puntuación y se le indican entonces al usuario un número predeterminado de aciertos, concretamente los que tienen la puntuación más alta. Si un usuario clica entonces sobre uno de los anuncios publicitarios, se incrementa su puntuación.
- La puntuación que indica la frecuencia con la que se ha clicado un anuncio publicitario puede emplearse también para calcular los costes que tiene que pagar el cliente publicitario al operador del motor de búsqueda. Aparte de la ordenación por rangos en base al número de clics sobre un anuncio, entran también en consideración otros factores que pueden influir sobre la ordenación por rangos. Así, por ejemplo, un cliente que pague más puede obtener en principio una "puntuación de bonificación".

40 Los resultados indicados pueden estar caracterizados por las propiedades siguientes:

- El URL que contiene la información
  - La indicación de la dirección que se ha asignado a la del URL
  - El vínculo con una representación de tarjeta de visita de las informaciones de firma
  - El vínculo con una representación de mapa del lugar asignado al URL
  - La representación puede efectuarse como texto corriente o como tabla
  - La representación de los resultados puede efectuarse como texto puro, con un mapa para orientación o con un mapa que represente todos los resultados.
- 45

La presente invención ha sido descrita con ayuda de varios ejemplos de realización. El experto entiende que la invención puede materializarse e implementarse programando un ordenador por medio de un lenguaje de programación usual de modo que éste esté en condiciones de ejecutar las funcionalidades de los ejemplos de realización descritos. Por consiguiente, el motor de búsqueda según ejemplos de realización de la invención consiste en un ordenador programado o bien en un programa informático que pone a un ordenador en condiciones de operar durante su ejecución con arreglo a las funcionalidades de los ejemplos de realización expuestos. Un procedimiento para realizar una funcionalidad de motor de búsqueda expuesta puede formar también un ejemplo de realización de la invención.

50

55

Según otro ejemplo de realización, el usuario puede ingresar las informaciones de origen geográfico no sólo, por ejemplo, por medio del ingreso de un lugar, sino también por medio del ingreso de un territorio, eventualmente también por selección sobre un mapa. Se recurre para ello nuevamente a los datos geocodificados a los que ya se

ha hecho referencia en el inventario de datos 230, de modo que se puede establecer aquí por medio de una imagen entre los datos geográficos y los nombres de lugar correspondientes el territorio geográfico al cual se refiere la petición de búsqueda.

5 Según otro ejemplo de realización, aparte de los contenidos de los propios aciertos se pueden indicar también otras informaciones adicionales para los aciertos localizados. A este fin, se puede indicar, por ejemplo, sobre una vista especialmente desglosada las informaciones adicionales de firmas asociadas, como, por ejemplo, año de fundación, productos, pero también dirección de correo electrónico, ID fiscal o bien el nombre de los gerentes. Estas se extraen también por medio de un mecanismo de regulación especial, análogamente al modo de proceder para la extracción de las informaciones locales.

10 A continuación, se describe un ejemplo de realización en el que se establece qué informaciones adicionales deberán indicarse todavía además de un acierto. Para una página localizada como acierto se establece aquí que términos se presentan con la mayor frecuencia en esta página. Se comprueba entonces para estos términos si coinciden con un término de la taxonomía que está almacenada en la base de datos 240 de la figura 2. Si ocurre esto, se puede partir entonces de la consideración de que este término presenta una cierta relevancia para la petición de búsqueda y, por  
 15 tanto, también para el usuario. El término correspondiente puede incorporarse después para el usuario sobre la propia página resultado o por separado asociado a la página resultado, concretamente en la forma en que se realiza por medio de un clic sobre este término una nueva petición de búsqueda sobre el inventario de datos que contiene este término de la taxonomía como criterio de búsqueda, o bien indicándole al usuario por medio de clicado sobre este término un número predefinido de vínculos predefinidos que caen bajo este término de la taxonomía. Estos  
 20 vínculos pueden consistir también en intercalaciones publicitarias que son intercaladas directamente por el operador del motor de búsqueda actuando como operador publicitario y que remiten a firmas que pagan al operador del motor de búsqueda por estas intercalaciones publicitarias. Los vínculos publicitarios están predefinidos y se encuentran sometidos a una ordenación por rangos según un ejemplo de realización, basándose la ordenación por rangos en la frecuencia con la que se ha clicado un enlace publicitario por un usuario. A este fin, se realiza para cada vínculo publicitario un recuento que cuenta el número de clics en este vínculo. Este recuento puede emplearse también  
 25 después para la facturación de los costes que carga en cuenta el operador del motor de búsqueda a quien haya encargado el anuncio.

30 Como alternativa al recuento de las palabras en un éxito en torno al tema que es relevante para este éxito y para el que deberán indicarse informaciones adicionales, se puede asignar también una entrada en el índice a un respectivo tema al que estén asignados nuevamente vínculos publicitarios. Estos vínculos publicitarios se indican (todos o los de rango más alto) cuando la entrada correspondiente del índice sea entregada como éxito a la petición de búsqueda.

**REIVINDICACIONES**

1. Motor de búsqueda para realizar una búsqueda por páginas de Internet para las que se cumple un criterio de origen geográfico ingresado por el usuario como término de búsqueda, presentando el motor de búsqueda:  
un dispositivo para escudriñar una pluralidad de páginas de Internet;
- 5 un dispositivo (135) para extraer datos de dirección de las paginas escudriñadas (130, 205), designando los datos de dirección extraídos una respectiva asignación geográfica de la página o del ofertante de la página;  
un dispositivo (140) para formar un inventario de datos (250) en el que está almacenado un gran número de páginas de Internet escudriñadas para cada una de las cuales se han extraído y asignado datos de dirección procedentes de estas páginas de Internet;
- 10 una interfaz de ingreso (145) para ingresar una petición de búsqueda por el usuario, posibilitándole la interfaz de ingreso al usuario el ingreso de un criterio de origen geográfico además de otros términos de búsqueda; y  
un dispositivo (150) para escudriñar el inventario de datos y para entregar las páginas de Internet para las cuales se cumplen el criterio de origen geográfico y los demás términos de búsqueda, por cotejo con los contenidos y los datos de dirección extraídos de las páginas de Internet;
- 15 **caracterizado** porque el dispositivo para extraer datos de dirección presenta:  
un dispositivo para aplicar un mecanismo de regulación al contenido de una página de Internet a fin de extraer las informaciones que pudieran especificar datos de dirección con arreglo a los criterios del mecanismo de regulación;  
un dispositivo para verificar los datos de dirección especificados con arreglo a los criterios del mecanismo de regulación por cotejo con un inventario de datos presente (220) de direcciones existentes, retirándose del inventario de datos (205) a escudriñar los datos de dirección especificados con arreglo a los criterios del mecanismo de regulación para los cuales el cotejo con el inventario de datos presente de direcciones existentes no suministra ninguna coincidencia.
- 20
2. Motor de búsqueda según la reivindicación 1, que presenta también:  
un dispositivo para dictaminar si la página escudriñada es la página de un ofertante comercial,
- 25
3. Motor de búsqueda según la reivindicación 1 ó 2, que presenta también:  
un dispositivo para geocodificar la página de Internet escudriñada estableciendo para una información de dirección extraída en base a los datos de dirección extraídos, por cotejo con un inventario de datos de geoinformación existentes, una geocodificación por medio de un sistema de coordenadas geográficas y asignando esta última a la página de Internet.
- 30
4. Motor de búsqueda según cualquiera de las reivindicaciones 1 a 3, que presenta también:  
un dispositivo para escudriñar las distintas páginas de Internet según una pluralidad de términos que son adecuados para clasificar las paginas de Internet atendiendo al contenido ofrecido, asignándose en el caso de un acierto - eventualmente aplicando otras condiciones - una clasificación correspondiente a la página de Internet.
5. Motor de búsqueda según cualquiera de las reivindicaciones 1 a 4, que presenta también:
- 35 un dispositivo para escudriñar por separado partes diferentes de las distintas páginas de Internet según la pluralidad de términos, para establecer y ponderar los aciertos en las diferentes partes y para establecer una puntuación total sobre la base de los aciertos ponderados, siendo decisivo en base a la puntuación total que se confiera una clasificación correspondiente a la respectiva página de Internet.
6. Motor de búsqueda según cualquiera de las reivindicaciones 1 a 5, que presenta también:
- 40 un dispositivo para indexar las páginas escudriñadas a las que se han asignado datos de dirección y eventualmente otras informaciones de clasificación;  
un dispositivo para emparejar los términos de búsqueda con el contenido del índice formado;  
una salida de los aciertos obtenidos, sirviendo los datos de dirección asignados a una página de Internet, por cotejo con el criterio de origen geográfico ingresado en calidad de término de búsqueda, como criterio de filtrado para la entrega de la lista de aciertos.
- 45
7. Motor de búsqueda según cualquiera de las reivindicaciones 1 a 6, que presenta también:

un dispositivo para formar varios índices parciales para partes diferentes de páginas web;

un dispositivo para emparejar los términos de búsqueda con el contenido de los respectivos índices parciales a fin de encontrar listas de aciertos parciales sobre la base de los índices parciales;

un dispositivo para formar una lista de aciertos totales por combinación de las listas de aciertos parciales.

5 8. Motor de búsqueda según la reivindicación 7, que presenta también:

una respectiva ponderación específica de aciertos de las diferentes listas de aciertos parciales; y

una combinación de las listas de aciertos parciales con las diferentes ponderaciones para formar una lista de aciertos totales sobre la base de las listas de aciertos parciales ponderadas combinadas.

9. Motor de búsqueda según la reivindicación 6, 7 u 8, que comprende también:

10 una asignación de una puntuación a un acierto individual de una lista de aciertos parciales sobre la base de un criterio de valoración para valorar la relevancia del acierto;

una asignación de una ponderación específica para la lista de aciertos parciales a un acierto individual de una lista de aciertos parciales;

15 un establecimiento de una puntuación total para el acierto de la lista de aciertos parciales por adición de las puntuaciones ponderadas que se han asignado a este acierto en las diferentes listas de aciertos parciales.

10. Motor de búsqueda según cualquiera de las reivindicaciones 1 a 9, que presenta también:

un dispositivo para la normalización de las informaciones geográficas extraídas a fin de poner éstas en un formato estandarizado que entrega después al usuario como "tarjeta de negocio" las informaciones de dirección y eventualmente otras informaciones de contacto junto con la dirección de Internet.

20 11. Motor de búsqueda según cualquiera de las reivindicaciones 1 a 10, que presenta también:

un inventario de datos de direcciones de Internet que deberán ser descargadas por la araña y escudriñadas por el dispositivo de extracción,

un dispositivo para la adaptación dinámica del inventario de datos por:

25 agregación de nuevos vínculos que se han establecido por el dispositivo de extracción al escudriñar las páginas descargadas; y/o

rechazo de direcciones de Internet en las que el dispositivo de extracción ha verificado que no se cumplen los criterios prefijados para la extracción de una información geográfica; y

un dispositivo para la descarga y escudriñamiento repetidos de las direcciones de Internet del inventario de datos.

12. Motor de búsqueda según cualquiera de las reivindicaciones 1 a 11, que presenta también:

30 un dispositivo para establecer informaciones adicionales a indicar que se indican además de los aciertos indicados ante una petición de búsqueda, presentando este dispositivo:

un dispositivo para establecer el sector temático para el que deberán indicarse informaciones adicionales.

13. Motor de búsqueda según la reivindicación 12, en el que el dispositivo para establecer el sector temático presenta:

35 un dispositivo para contar la frecuencia de palabras individuales que se presentan en el acierto o aciertos a fin de establecer en base a las palabras que se presentan con la mayor frecuencia el sector temático para el cual deberán indicarse informaciones adicionales: y/o

un dispositivo para consultar en sectores temáticos asignados a los respectivos aciertos a fin de establecer basándose en ellos el sector o los sectores temáticos para los cuales deberán indicarse informaciones adicionales.

40 14. Motor de búsqueda según la reivindicación 12 ó 13, en el que las informaciones adicionales indicadas son vínculos publicitarios y la secuencia de indicación de los vínculos publicitarios se basa en la frecuencia con que se haya clicado ya en un vínculo publicitario.

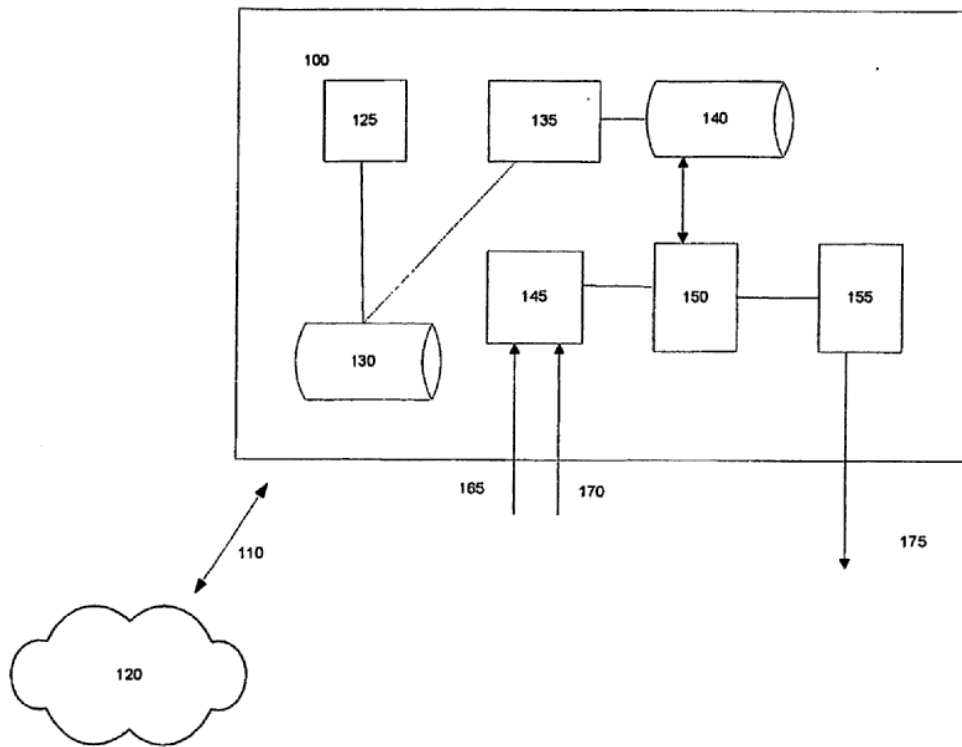


Fig. 1

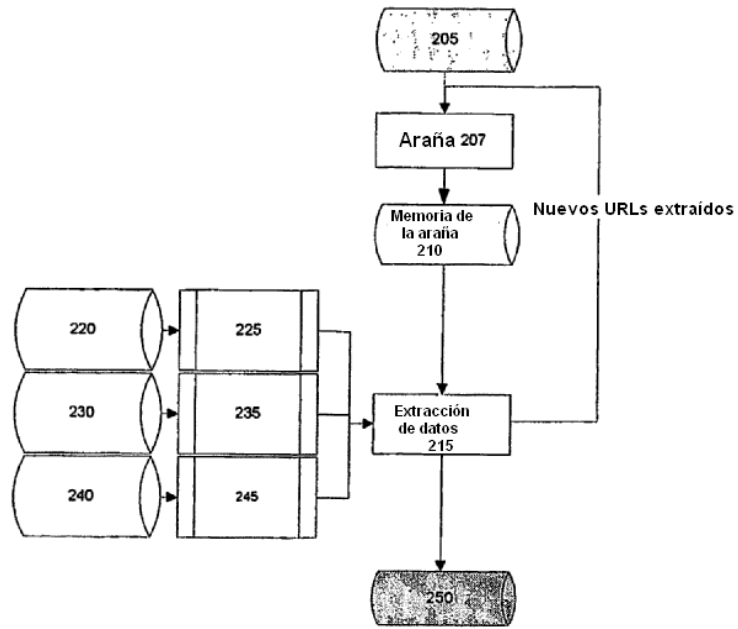


Fig. 2



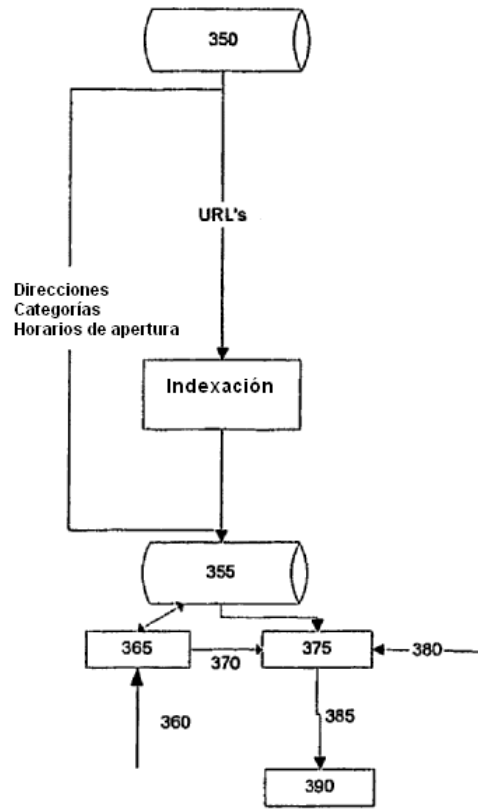


Fig. 3