(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 329 008**

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

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **01904831 .3**
(96) Fecha de presentación : **12.01.2001**
(97) Número de publicación de la solicitud: **1381967**
(97) Fecha de publicación de la solicitud: **21.01.2004**

(54) Título: **Red de información distribuida accesible de forma global.**

(30) Prioridad: **14.01.2000 US 176329 P**

(45) Fecha de publicación de la mención BOPI:
20.11.2009

(45) Fecha de la publicación del folleto de la patente:
20.11.2009

(73) Titular/es: **THINKSTREAM, Inc.**
9386 Ashland Road
Gonzalez, Louisiana 70737, US

(72) Inventor/es: **Malone, Michael, K.;**
Loucks, Jon, David;
Terhune, Robert, K., III y
Cotrill, Cameron, W.

(74) Agente: **Carpintero López, Mario**

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).

Venta de fascículos: Oficina Española de Patentes y Marcas. Pº de la Castellana, 75 – 28071 Madrid

DESCRIPCIÓN

Red de información distribuida accesible de forma global.

Campo técnico

La presente invención se refiere a sistemas y técnicas para recopilar y buscar información disponible en sitios de una red de información accesible de forma global tal como Internet y, en particular, a una arquitectura de búsqueda distribuida que facilita el acceso en tiempo real a información que reside en cualquier número de servidores distribuidos por toda la red y sintetiza la información para un acceso continuo a información específica buscada por un usuario.

Antecedentes de la invención

Aunque ha demostrado un crecimiento explosivo y ha tenido un enorme impacto en los ámbitos de la información y el comercio, la red informática accesible de forma global, conocida como Internet, se ha convertido, en efecto, en una víctima desestructurada de sí misma. La utilización de información de Internet ha perdido en gran medida su utilidad debido a que los motores de búsqueda tradicionales no pueden ni acceder al vasto fondo de información disponible ni calificarlo de forma adecuada. El mejor motor de búsqueda actual puede realizar un seguimiento de, y acceder a, sólo una pequeña fracción de las páginas de la World Wide Web de Internet (es decir, aproximadamente mil millones de 550 mil millones de documentos disponibles). Los sitios accesibles se dividen en categorías de forma rudimentaria usando palabras clave en vez de una valoración inteligente del contenido. Un resultado común actual de búsquedas de información, incluso limitadas a la pequeña fracción de la información disponible, es de miles, y a menudo millones, de respuestas irrelevantes.

La recopilación y distribución de información en Internet tiene lugar como sigue. Un motor de búsqueda por Internet convencional usa software (denominado "*spiders*") que itinera por la Web para recopilar información, que se destila, indexa y cataloga en una base de datos central. Una búsqueda por Internet realizada por un usuario Web de ese motor de búsqueda produce resultados que provienen de la base de datos, no de la propia Internet. Los resultados producidos son referencias a direcciones de Internet, requiriendo de este modo que el usuario Web abra múltiples sitios en la búsqueda de la información.

Los motores de búsqueda actuales no incluyen una capacidad de búsqueda masiva de todos los sitios y recuperan y organizan los resultados de búsqueda por contenido; por tanto, las búsquedas se aplican a toda la información accesible, con independencia de si es relevante. El resultado es un esfuerzo de motor de búsqueda en gran medida ineficaz y respuestas no sensibles a las consultas de búsqueda. Ejemplos de tales motores de búsqueda tradicionales incluyen Northern Light™, Snap™, Alta Vista™, HotBot™, Microsoft™, Infoseek™, Goggle™, Yahoo™, Excite™, Lycos™ y Euroseek™.

La tecnología de búsqueda convencional se basa, por tanto, en un modelo en el que los índices, referencias y datos reales (en el caso de redes de comercio) están centralizados. Todas las consultas tienen lugar en sitios centrales, y los datos distribuidos no se actualizan en tiempo real (y están normalmente obsoletos) y normalmente requieren un nuevo formateo. Internet es, en el mejor de los casos, un entorno de búsqueda frustrante, debido a que los datos residen en múltiples formatos y en un ámbito distribuido.

Para aplicaciones de comercio, la arquitectura de Internet existente puede albergar sólo a una pequeña fracción de la participación empresarial que en otro caso estaría disponible para producir beneficios para el consumidor derivados de la competencia. Internet, como consecuencia, sirve en efecto sólo a los grandes participantes dominantes, mientras que excluye efectivamente a todos los demás. Parte de la percepción del comercio electrónico es que puede comprarse virtualmente cualquier cosa por Internet. Aunque la percepción es precisa, ignora el hecho de que la polarización en el sistema actual deja fuera una parte del mercado mucho mayor que a la que da servicio. La utilización comercial de empresa a empresa de Internet consiste en gran medida en comunicaciones por correo electrónico.

Para aplicaciones en prestación de servicios, particularmente como diversas entidades gubernamentales han intentado usar Internet, la carencia de estructura práctica es especialmente notable. Estas situaciones no existen por culpa o incompetencia de los usuarios, sino que de nuevo tienen su origen en una limitación inherente y sistémica de una Internet "centralizada".

Los esfuerzos de sitios de búsqueda tradicionales para conservar y atraer más atención del consumidor y de este modo generar más ingresos por publicidad han provocado que el intento de centralizar toda la información en línea llegue a un punto de conflicto. Según se expone anteriormente, el crecimiento en el volumen y la diversidad de contenido de Internet en la actualidad conduce a búsquedas que generan miles de páginas de resultados que abarcan sólo una fracción del cuerpo global de la información relevante. El mercado necesita acceder a estructuras organizativas adicionales, pero el sistema actual hace estos requisitos imposibles de cumplir. Los sitios de búsqueda tradicionales se diseñan y predicen para conducir a una centralización adicional, que acentuará el problema de la accesibilidad a la información.

La idea convencional ha sido que la velocidad puede compensar el crecimiento de la información de Internet. El énfasis del sector se ha puesto en mejoras de hardware en vez de en software de próxima generación. Hace cinco años,

un ordenador personal según el estado de la técnica usaba un chip microprocesador de 166 MHZ. Actualmente, chips microprocesadores de 800 MHZ son convencionales, y se espera que chips microprocesadores de 1.000 MHZ estén disponibles pronto. Irónicamente, aunque las máquinas disponibles actualmente pueden buscar información mucho más rápidamente, también crean información a un ritmo acorde con su velocidad. Están, en efecto, ayudando a que el problema siga el ritmo de la solución. En la medida en que el énfasis se ha puesto en el software, ha sido para mejorar las aplicaciones dentro de la arquitectura actual o para ofrecer y comercializar alternativas al comercio electrónico dentro de la arquitectura actual. Como consecuencia, se ponen trabas a todos los esfuerzos de este tipo antes de que comiencen.

Debido al mero tamaño de Internet y a que los *spiders* operan desde una ubicación central, los *spiders* pueden cubrir sólo una pequeña fracción de todo Internet. La base de datos resultante de resultados de búsqueda está inherentemente limitada no sólo en tamaño sino también en actualidad. Las compensaciones requeridas son contraproducentes. Hacer la base de datos más amplia y más profunda requeriría un tiempo de "itinerancia" excesivo de modo que la información se volvería obsoleta. Mantener la información actual requeriría buscar en una fracción más pequeña de los documentos de Internet disponibles, haciendo de este modo los resultados menos exhaustivos.

La información total está creciendo en la actualidad a un ritmo exponencial. La mayor parte de la información nueva termina en la categoría de inaccesible. No hay certeza de que la información actualizada "echará" a la información desfasada del fondo de información accesible. La edad promedio de los enlaces de la World Wide Web recientemente devueltos es de 186 días. El entorno es frecuentemente el de información antigua, información insuficiente, información desorganizada y, en resumen, información no gestionable. Existe una necesidad acuciante, por tanto, de incorporar la Internet existente en un nuevo ámbito de organización eficaz que gestione, de forma competente, generaciones futuras de crecimiento.

El documento de HINDS N *ET AL*: "Managing metadata for distributed information servers" SYSTEM SCIENCES, 1998., *PROCEEDINGS OF THE THIRTY-FIRST HAWAII INTERNACIONAL CONFERENCE ON KOHALA COAST*, HI, EE.UU. 6-9 de enero de 1998, LOS ALAMITOS, CA, EE.UU., IEEE COMPUT. SOC.US, 6 enero de 1998, páginas 513 a 522, se refiere a una arquitectura de índice distribuido para facilitar encontrar metadatos en grandes redes, que usa jerarquías descriptivas para organizar metadatos y encaminar consultas de metadatos a sitios que es probable que produzcan resultados relevantes.

Sumario de la invención

Aspectos de la invención se exponen en las reivindicaciones adjuntas. La presente invención es una red de información distribuida que se construye para recopilar información a partir de sitios distribuidos a través de una red informática accesible de forma global, es decir, Internet. Estos sitios distribuidos están equipados para alojar y mantener su propia información, mientras que otra tecnología asociada habilita la inclusión de sitios individuales en las búsquedas masivas por Internet.

Una realización preferida de la red de información distribuida incluye un servidor raíz que almacena una lista de múltiples sitios distribuidos, cada uno de los cuales está representado por metadatos correspondientes a contenido de información disponible directa o indirectamente. Los metadatos son propiedades ampliadas de un objeto de datos, que pueden ser, por ejemplo, un único archivo, un objeto en una base de datos, un mensaje de correo electrónico, un fragmento de memoria o una descripción del contenido de información en un sitio. Los metadatos pueden ser tan simples como para representar un nombre o tamaño de archivo o tan complejos como para representar al autor del archivo o información de esquema de base de datos. El explorador de red del usuario entrega una petición de búsqueda de información al servidor raíz, que en respuesta desarrolla una petición de búsqueda de información perfilada. Cada uno de múltiples sitios distribuidos se implementa con un proveedor de información que está ubicado de forma remota del servidor raíz. El proveedor de información de cada uno de los sitios distribuidos almacena metadatos correspondientes a contenido de información que puede recuperarse en respuesta a la petición de búsqueda de información perfilada de resultados de búsqueda que pueden derivarse del contenido de información al que corresponden los metadatos. Un enlace de comunicación de información perfilada entre el servidor raíz y cada uno de los múltiples sitios de distribución habilita la formación de una trayectoria para la entrega de los resultados de búsqueda a un sitio de destino, tal como el explorador de red, a partir de un sitio o sitios representados por los metadatos de la petición de búsqueda de información perfilada.

La realización preferida descrita anteriormente de una red de información distribuida proporciona un motor de búsqueda por Internet que usa de forma ventajosa los puntos fuertes inherentes de Internet: una arquitectura distribuida. Cuando se inicia una petición de búsqueda, el motor de búsqueda consulta a múltiples sitios de forma simultánea y busca la información, en cualquier formato de datos en que ésta resida, encuentra la información, y entonces devuelve el documento real al usuario. Un explorador web cliente habilitado para multihilo envía consultas simultáneas a servidores distribuidos, eliminando de este modo el cuello de botella de un cuerpo de búsqueda o servidor centralizado. El explorador web cliente también gestiona la descarga de información desde el servidor y, por tanto, lo habilita para manejar un número drásticamente mayor de clientes que los manejados por modelos tradicionales de la actualidad. Esta aplicación de búsqueda distribuida trata las deficiencias fundamentales en la cobertura de Internet actual: acceso pobre, almacenes de datos obsoletos, información irrelevante, y repositorios no estructurados de información infrautilizada.

La arquitectura de búsqueda de la invención incluye la capacidad de realizar una búsqueda descentralizada de datos en directo (estructurada o no estructurada), búsqueda según parámetros específicos (precio, marca, disponibilidad, reseñas y otros parámetros de ese tipo), y presentar los resultados de búsqueda de forma limpia y organizada en una pantalla de visualización. La arquitectura de búsqueda en efecto mueve la consulta a la ubicación de la información. Un usuario puede aplicar de forma continua filtros a los resultados de búsqueda y centrarse en el producto o información específico que está buscando el usuario.

Las ventajas de la arquitectura de búsqueda distribuida incluyen conformidad con las normas del sector; escalabilidad vertical y horizontal; sin requisitos de hardware adicional o empeoramiento del rendimiento; uso de ancho de banda disponible de Internet en lugar del ancho de banda disponible de cualquier motor de búsqueda central, eliminando de este modo posibles cuellos de botella inherentes a cualquier solución centralizada; entrega de información precisa y actual; requisito de recursos de infraestructura inferiores (servidores, almacenamiento electrónico y ancho de banda), como consecuencia de estar las consultas distribuidas por toda la red; ausencia de empeoramiento del rendimiento en relación con el número de sitios buscados y ausencia de limitaciones impuestas al número de sitios buscados; ausencia de efecto de sitios caídos en los resultados de búsqueda; y gestión de cliente de toda la clasificación, filtrado y comparaciones de datos, eliminando de este modo el tráfico de red y el procesamiento de datos redundante actualmente requerido por las arquitecturas de hoy en día.

El uso de sitios distribuidos representa un cambio fundamental con respecto al actual procedimiento de almacenamiento masivo central y abre las puertas a la gran fracción restante de información almacenada, aunque inaccesible con la arquitectura actual. El resultado es la creación de vastas áreas de nuevas oportunidades en el comercio electrónico e intercambio de información corporativa a través de portales de información. Tales nuevas oportunidades incluyen aplicaciones en la distribución de música y películas, distribución de aplicaciones software, mensajería instantánea, colaboración, subastas, comercio individual, búsquedas paralelas y correo electrónico. Este cambio permite una interacción de comercio electrónico de consumidor y de empresa a empresa (B2B) más sofisticada.

La presente invención proporciona una oportunidad para establecer nuevas normas y procedimientos para recopilar información a partir de sitios distribuidos a través de Internet. La invención está adaptada para seguir el ritmo del crecimiento actual de la World Wide Web y puede aplicarse a prácticamente cada comerciante, corporación y consumidor. Los sitios distribuidos pueden alojar y mantener su propia información mientras la invención permite que los sitios individuales se incluyan en las búsquedas masivas por Internet. La invención se implementa como una única arquitectura distribuida, con su propio motor de búsqueda inteligente, para gestionar información digital y usa software para la gestión de Internet y de su contenido, para lograr resultados sensibles a partir de búsquedas por Internet.

La arquitectura distribuida puede describirse de forma análoga, conceptualmente, como similar a prefijos telefónicos o códigos postales. La diferencia es que la codificación es específica del contenido en vez de específica de la zona geográfica. La arquitectura de red de información distribuida puede buscar sitios existentes, incluyendo el 84% de sitios inaccesibles actualmente, dividirlos en categorías de forma inteligente según el contenido y codificarlos según se requiera con códigos únicos o múltiples para su futura recuperación inteligente. Los futuros sitios pueden integrarse fácilmente a medida que se ponen en línea para estar disponibles de forma inmediata, poniendo fin de este modo al retardo actual de 186 días. Si se desea, los usuarios de comercio pueden descargar software de sitio web de comercio electrónico que permite una presentación personalizada de todo el inventario de productos ofrecido. Un cliente que quiera comprar un producto particular puede comparar de forma inmediata, a través de múltiples sitios de vendedor, por ejemplo, precios de vendedor, garantías, políticas de devolución y gastos de envío.

La tecnología y red de búsqueda distribuida puede aplicarse a comercio electrónico y sirve para eliminar la polarización, dando como resultado de este modo que se da servicio al comercio pequeño y tradicional al igual que a los superalmacenes electrónicos que actualmente dominan la oferta de productos y servicios. Los vendedores pequeños y tradicionales tienen pocas oportunidades de crear visibilidad dentro de los confines del mercado actual, debido a que los resultados de búsqueda se comercializan y no existe provisión para comparaciones de producto “en directo” reales. La invención presenta una oportunidad sustancial para buscar resultados que conduzcan a un producto real, en vez de a un sitio web, y ofrece de este modo soluciones que eliminan la polarización y conducen a un entorno de competencia leal en el que puede garantizarse a los vendedores que sus sitios y productos están incluidos.

La invención permite que vendedores y corporaciones ejerzan el control sobre los tiempos y el contexto de su propia información y facilita una tendencia a la “descentralización” como una etapa evolutiva natural para Internet. El motor de búsqueda también funciona dentro de un portal de información que permitirá cooperación B2B eficaz. Por ejemplo, los vendedores de componentes ya no requieren enlaces de sistema directos con los fabricantes de equipo originales para asegurar un suministro oportuno y adecuado. La invención permite la selección inmediata de categoría, línea de producto y nombre de marca. Todos los vendedores registrados en la arquitectura están representados para su comparación. La invención hace posible que existan mercados sustancialmente verticales para sus soluciones en las que redes privadas de información estructurada y que puede buscarse pueden usarse para crear sistemas de suministro y adquisición y redes de indagación de información.

Breve descripción de los dibujos

La figura 1 es un diagrama de bloques de un ejemplo de una red de aplicación distribuida configurada según la presente invención.

La figura 2 es un diagrama de bloques que muestra más detalladamente la estructura interna del servidor raíz mostrado en la figura 1.

La figura 3 es un diagrama de bloques de un servidor de sitio de nivel uno, que muestra el flujo de programa cuando se realiza una consulta distribuida en la red de aplicación distribuida de la figura 1.

La figura 4 es un diagrama de bloques de un servidor de nodo de sitio de nivel dos que no tiene sitios registrados con el proveedor de sitio y no tiene ningún servidor hijo.

La figura 5 es un diagrama de bloques de un servidor de sitio en el que coexisten varios proveedores diferentes para una amplia variedad de fuentes de información.

La figura 6 es un diagrama de bloques que muestra un gestor de analizadores de servidores de sitio y sus analizadores para un elemento de acceso a archivos y su almacenes de datos para su uso para respaldar una explicación de un procedimiento de acceso y análisis de datos según una realización de la invención.

La figura 7 es un diagrama de bloques que muestra en mayor detalle la estructura y organización de determinados bloques de componente de la figura 6.

La figura 8 es un diagrama de bloques de una red de información distribuida compuesta por una red de comercio electrónico, una red de empresa a empresa, una red del lado de suministro de empresa a empresa y una red de información implementada con servidores públicos y privados.

La figura 9 es un diagrama de bloques que muestra más detalladamente la estructura interna de un grupo ovoide de aplicación de información de la red de información distribuida de la figura 8.

La figura 10 es un diagrama de flujo de un proceso de seguridad y de autenticación de sesión para comunicaciones de red entre pares según una realización de la invención.

La figura 11 es un diagrama de flujo que define en términos generales las etapas de un proceso para proporcionar seguridad para compartir archivos en un entorno distribuido.

Descripción detallada de realizaciones preferidas

La figura 1 es un diagrama de bloques de un ejemplo de una red 10 de aplicación distribuida configurada según la invención y que muestra trayectorias de flujo de información en respuesta a una petición de usuario final particular. Una red de aplicación es un agrupamiento de servidores que participan en una aplicación particular de la red de información distribuida de la invención. Ejemplos de una red de aplicación incluyen una red de comercio electrónico, un portal de información o una red entre pares (P2P). La red 10 es un sistema jerárquico de servidores distribuidos que almacenan el contenido de red y se comunican con otros servidores en la red. El sistema jerárquico es uno en el que un servidor puede tener cualquier número de servidores hijos, cada uno de los cuales puede tener cualquier número de sus propios servidores hijos, con un número ilimitado de niveles sucesivos de servidores dependientes posibles. Esta estructura ayuda a distribuir el almacenamiento del contenido y la carga de procesamiento en la red. Las figuras 2 a 4 muestran más detalladamente las estructuras internas de, respectivamente, servidores raíz, de sitio y de nodo de sitio representados como bloques de componente de sistema en la figura 1. Las figuras 1 a 4 soportan la siguiente visión general explicativa de la tecnología principal implementada en una arquitectura distribuida de Internet que opera en respuesta a una búsqueda típica de contenido por un usuario.

Con referencia a la figura 1, la red 10 incluye un cliente de sistema operativo, que es normalmente un explorador web o *applet* 12 cliente que se almacena en un ordenador de usuario final. El *applet* cliente es software del lado de cliente que está escrito preferentemente en código de lenguaje JAVA (aunque pueden escribirse en cualquier otro lenguaje de desarrollo de software) y permite a cualquier ordenador participar en la red. El *applet* 12 cliente es la interfaz de software entre el usuario y la red de aplicación. Un servidor 14 raíz ubicado de forma remota del ordenador del usuario se implementa con un perfilador raíz que almacena una lista de múltiples sitios distribuidos a través de una red informática global, tal como Internet. El servidor 14 raíz es el único "antepasado" de todos los servidores y servidores hijos y es el punto de entrada principal para el *applet* 12 cliente. El servidor 14 raíz tiene tres hijos, los servidores 16, 18 y 20 de sitio, que representan servidores de nivel uno de la Compañía A, la Compañía B y la Compañía C, respectivamente. Los servidores 16, 18 y 20 de sitio representan ejemplos de fuentes de información enumeradas en el perfilador raíz del servidor 14 raíz y válidas en respuesta a una petición específica del usuario. Los expertos apreciarán que existen numerosas fuentes de información candidatas diferentes, tales como, por ejemplo, redes estatales y otras redes gubernamentales, datos corporativos, sitios web de información comercial y educativa, sitios web de comercio electrónico y ordenadores personales individuales de sobremesa (PC).

Cada uno de los servidores 16, 18 y 20 de sitio se implementa con un proveedor de información que almacena metadatos que pueden recuperarse, que se mantiene actual por y bajo el control de la compañía con la que está asociado el servidor de sitio. Los metadatos son información acerca del contenido residente de forma local almacenado en cada servidor de sitio y el contenido en cualquier servidor hijo que pueda tener un servidor de sitio. Existen dos tipos básicos de metadatos, que son datos de tema y datos de perfil de sitio. Un tema es una unidad de contenido ofrecida

por una red de aplicación. La base de datos de tema en un servidor de sitio almacena información acerca del tipo de información almacenada en el sitio y sus sitios hijos. (En las figuras 2 y 3, las bases de datos de tema están etiquetadas, respectivamente, "Base de datos de tema" en el servidor 14 raíz y bases de datos de "Tipo de contenido" en el servidor 16 de sitio). La base de datos de perfil de sitio almacena información acerca de cuáles de los servidores, incluyéndose a sí mismo y sus hijos, almacenan qué tipos de temas. Los servidores 16, 18 y 20 de sitio proporcionan, por tanto, un conjunto de bases de metadatos, que son bases de datos de información acerca de la información que se almacena e intercambia en la red 10 y que son bases de datos que realizan un seguimiento de dónde se almacenan tipos particulares de información en la red 10. El perfilador raíz identifica los servidores 16, 18 y 20 de sitio mediante códigos específicos de contenido que representan perfiles de tema indicativos del contenido de información que los servidores 16, 18 y 20 de sitio contienen. El servidor 16 de sitio de la Compañía A está asociado con un servidor de nivel dos, el servidor 22 de nodo de Sitio A. El servidor 20 de sitio de la Compañía C está asociado con dos servidores de nivel dos, el servidor 24 de nodo de Sitio C y el servidor 26 hijo de Sitio C. El servidor 26 hijo de Sitio C está asociado con dos servidores de nivel tres, el servidor 28 de nodo de Sitio C2 y el servidor 30 de nodo de Sitio C2.

La figura 1 ilustra la operación de la red 10 cuando un usuario hace que el explorador 12 web solicite al servidor 14 raíz la identificación de servidores válidos relacionados con un tema específico. El *applet* 12 cliente envía la petición a los servidores 16, 18 y 20 de sitio, todos los que el servidor 14 raíz ha identificado como válidos en respuesta al tema que solicitó el usuario. (Las líneas discontinuas con punta de flecha dibujadas entre el servidor 14 raíz y cada uno de los servidores 16, 18 y 20 de sitio representan trayectorias de comunicación para actualizar metadatos acerca de sitios en la red y actividad de relación (por ejemplo, información y seguimiento de transacciones) que los enlaza y no indican trayectorias de búsqueda).

La red 10 procesa una petición de consulta de tema de usuario como sigue. Un usuario de red explora una página web en el servidor 14 raíz. Si no está ya instalado en el ordenador personal del usuario, el *applet* cliente se descarga e instala (con el permiso del usuario). El *applet* 12 cliente descarga una base 48 de datos de tema actual del servidor 14 raíz, que muestra la estructura de temas normalmente como un árbol jerárquico de categorías. El *applet* 12 cliente permite entonces al usuario navegar por el árbol de categorías hasta que el usuario encuentra la categoría de temas de interés. Tan pronto como el usuario navega a un nivel de categoría que tiene la suficiente especificidad para estar asociado con servidores de sitio particulares, el *applet* 12 cliente envía una consulta o bien automática o bien controlada por el usuario al servidor 14 raíz. Cuando el *applet* 12 cliente indica una búsqueda, se envía al servidor 14 raíz la petición de consulta de una lista de servidores de sitio válidos. El servidor 14 raíz devuelve al *applet* 12 cliente un paquete de información que contiene una lista de todos los servidores de sitio válidos en la red 10 de aplicación que tienen el tipo del contenido solicitado. Los servidores 16, 18 y 20 de sitio representan los servidores de sitio que aparecen en la lista en el ejemplo ilustrado en la figura 1. A medida que el usuario navega por el árbol en sentido descendente hacia el nivel de tema, el *applet* 12 cliente usa los metadatos disponibles para visualizar un selector de atributos. Esto permite que el usuario seleccione atributos, características, rasgos, especificaciones y otros aspectos especificados del tema, que permiten al usuario concretar el objetivo de la búsqueda. Cuando la consulta de tema es lo suficientemente específica, el usuario la ejecuta. El *applet* 12 cliente del usuario en este ejemplo compila una lista de servidores 16, 18 y 20 de sitio, realiza una consulta de tema en cada uno de ellos y espera a los resultados que producen los servidores 16, 18 y 20 de sitio. El procesamiento de la petición de consulta de tema conlleva dirigirla a los tres servidores 16, 18 y 20 de sitio de nivel uno. Los servidores 16 y 20 de sitio pasan entonces la petición de consulta de tema a los tres servidores 22, 24 y 26 de nivel dos. El servidor 26 hijo de Sitio C pasa además la petición de consulta de tema a los servidores 28 y 30 de nodo de Sitio C2. Este proceso tiene lugar al tiempo que se evita cualquier servidor que no tenga el contenido pertinente. Los resultados obtenidos se devuelven, al tiempo que se evitan de nuevo todos los demás servidores, al *applet* 12 cliente, para su visualización al usuario. El usuario puede entonces evaluar los resultados de búsqueda y hacer clic sobre cualquiera de las fuentes de contenido enlazadas. El software 32 de aplicación de administración (figuras 2 y 3) se comunica con el servidor 14 raíz para realizar un seguimiento del número y tipos de peticiones de búsqueda de tema procesadas, así como para actualizar las bases de metadatos en los servidores de sitio.

La figura 2 es un diagrama de bloques que muestra en mayor detalle la estructura interna del servidor 14 raíz. La figura 2 muestra el flujo de programa cuando una lista de servidor de sitio se compila en el servidor 14 raíz y se entrega al *applet* 12 cliente en respuesta a una petición de consulta de tema realizada por un usuario. Con referencia a la figura 2, la petición de consulta de tema iniciada por el *applet* 12 cliente pasa a través de la World Wide Web a un servidor 50 web en el que se almacenan páginas web asociadas con el servidor 14 raíz. (El servidor 50 web puede estar separado físicamente o ser una parte del servidor 14 raíz). El servidor 50 web pasa la petición de consulta de tema al servidor 14 raíz, que usa sus proveedores de información para consultar en su base de datos todos los servidores que corresponden al tipo de la petición. El servidor 14 raíz se implementa con una interfaz 52 de analizador de consulta que incluye un proveedor 54 de sitio y un proveedor 56 principal para interpretar la petición de consulta de tema. Cada uno del proveedor 54 de sitio y el proveedor 56 principal es preferentemente un programa basado en lenguaje JAVA que se ejecuta en el servidor 14 raíz. Los componentes de proveedor 54 de sitio y proveedor 56 principal de la interfaz 52 de analizador de consulta consultan las bases de metadatos locales para determinar qué servidores de sitio conducen al tipo específico de contenido de tema solicitado. Esto conlleva identificar servidores de sitio que tienen en sí mismos los temas correctos o están asociados con servidores descendientes que tienen los temas correctos. El proveedor 54 de sitio identifica servidores de sitio correspondientes a los códigos específicos de contenido que representan los perfiles de tema, y el proveedor 56 principal identifica propiedades de los temas. La interfaz 52 de analizador de consulta accede y recupera información de la base 48 de datos de tema y una base 60 de datos de perfil de sitio para reunir el paquete de información que contiene la lista de servidores de sitio válidos en los que va a buscarse. El paquete de información representa una petición de búsqueda de información perfilada generada por el servidor 14 raíz. Un

módulo 62 de interfaz administrativa contiene software para mantener las bases de datos y facilitar información acerca de la frecuencia de acceso a las mismas.

Un ejemplo de una petición de consulta de tema es la identificación de vendedores de VCR de un tipo particular. El proveedor 54 de sitio recupera a partir de la base 60 de datos de perfil de sitio las identidades de servidores de sitio de compañías que venden VCR. El proveedor 56 principal recupera a partir de la base 48 de datos de tema las propiedades (por ejemplo, coste de compra, compatibilidad con disco compacto y capacidad de sonido estereofónico) del tipo especificado de VCR. El servidor 14 raíz devuelve el paquete de información reunido al usuario por medio del servidor 50 web. La petición de consulta de tema se distribuye entonces a través del *applet* 12 cliente a los servidores de nivel uno de los sitios identificados.

La figura 3 es un diagrama de bloques del servidor 16 de sitio de nivel uno, que muestra el flujo de programa cuando se realiza una consulta de tema solicitada. (Aunque el servidor 16 de sitio tiene sólo el servidor 22 de nodo, la figura 3 muestra en líneas discontinuas dos servidores de sitio hijo de mayor nivel jerárquico para demostrar la escalabilidad de la red). Con referencia a la figura 3, el servidor 16 de sitio recibe del *applet* 12 cliente una petición de consulta de tema realizada por un usuario y perfilada por el servidor 14 raíz. El servidor 16 de sitio se implementa con una interfaz 78 de analizador de consulta y procesa la petición de consulta de tema determinando si el propio servidor 16 de sitio o un servidor de sitio de nodo hijo asociado puede soportar la consulta de tema. La interfaz 78 de analizador de consulta incluye un proveedor 82 de sitio, un proveedor 82 de tipo de contenido A, un proveedor 84 de tipo de contenido B y un proveedor 86 de tipo de contenido C, todos los cuales representan diferentes formas de recopilar información de contenido conectando una petición de consulta de tema y una base de datos. Por ejemplo, los tipos A, B y C de contenido pueden representar, respectivamente, información de comercio electrónico, datos y contenido de sitio (HTML).

El proveedor 80 de sitio, el proveedor 82 de comercio electrónico, el proveedor 84 de datos y el proveedor 86 de HTML acceden y recuperan información de contenido de, respectivamente, una base 90 de datos de sitio hijo, una base 92 de datos de tipo de contenido A (de comercio electrónico), una base 94 de datos de tipo de contenido B (datos), y una base 96 de datos de tipo de contenido C (contenido de sitio (HTML)). Cada servidor de sitio de nodo hijo devuelve sus resultados de búsqueda al servidor 16, como se describe a continuación con referencia a la figura 4. Los proveedores de información de la interfaz 78 de analizador de consulta y los resultados de búsqueda recibidos desde cualquier sitio de nodo hijo son las fuentes a partir de las cuales el servidor 16 de sitio construye una lista de sitio que devuelve los resultados completos de la búsqueda al *applet* 12 cliente.

Cuando el contenido en cualquier servidor cambia, un administrador de sitio usa software 32 de aplicación de administración (figuras 2 y 3) para actualizar las bases de metadatos en el servidor de sitio. Esas actualizaciones se envían automáticamente a todos los servidores padre asociados de mayores niveles jerárquicos. Una interfaz de administración de cada servidor (interfaz 98 administrativa del servidor 16) en cada nivel (y la interfaz 62 administrativa del servidor 14 raíz) actualiza las bases de metadatos locales. Cada servidor a lo largo de un linaje siempre tiene una representación actual del contenido disponible de forma local y a través de sus sitios hijos. El servidor 14 raíz aloja, por tanto, bases de metadatos completas y actuales de qué tipo de información se almacena en la red 10 (en la base 48 de datos de tema) y la primera etapa en la trayectoria hacia donde se almacena la información en la red 10 (en la base 60 de datos de perfil de sitio).

La figura 4 es un diagrama de bloques de un servidor 22 de nodo de Sitio A de nivel dos, que no tiene ningún sitio registrado con su proveedor 100 de sitio y no tiene servidor hijo. Con referencia a la figura 4, un proveedor 102 de tipo de contenido A (comercio electrónico), un proveedor 104 de tipo de contenido B (datos) y un proveedor 106 de tipo de contenido C (HTML) que residen en la interfaz 108 de analizador de consulta del servidor 22 de nodo de Sitio A proporcionan temas válidos para buscar en una base 110 de datos de tipo de contenido A (de comercio electrónico) y una 112 base de datos de tipo de contenido B (sitio). Los resultados obtenidos de las búsquedas de las bases 100 y 102 de datos se devuelven al servidor 16 de sitio padre para la entrega al *applet* 12 cliente. Una interfaz 114 administrativa actualiza las bases de metadatos locales.

El servidor 16 de sitio, junto con el servidor 22 de nodo de Sitio A; el servidor 20 de sitio, junto con el servidor 24 de nodo de Sitio C; y el servidor 20 de sitio, junto con el servidor 26 hijo de Sitio C y el nodo 30 de sitio C2, forman cada uno una red de información local según la invención.

El servidor 16 de sitio puede implementarse con un perfilador raíz local, que como se indica en la figura 1, incluye el servidor 22 de nodo de Sitio A en su lista de sitios locales distribuidos. El servidor 22 de nodo de Sitio A también es ampliable para alojar su propio perfilador raíz local aunque en el ejemplo representado en las figuras 1 y 4 proporciona sólo metadatos locales en respuesta a una petición de búsqueda de información perfilada local acompañada por un código local específico del contenido de información correspondiente al contenido de información de los metadatos locales.

El servidor 20 de sitio puede implementarse con un perfilador raíz local, que como se indica en la figura 1, incluye el servidor 24 de nodo de Sitio C y el servidor 26 hijo de Sitio C en su lista de sitios locales distribuidos. De manera similar, el servidor 26 hijo de Sitio C puede implementarse con su propio proveedor raíz local, que como se indica en la figura 1, incluye los servidores 28 y 30 de nodo de Sitio C2 en su lista de sitios locales distribuidos. Cada uno de los nodos 28 y 30 de Sitio C2 es también ampliable para alojar su propio perfilador raíz local.

Los sitios incluidos en los servidores de nivel uno y servidores en niveles sucesivos funcionan, por tanto, o bien para enumerar sitios distribuidos o bien para proporcionar metadatos para su procesamiento por la red distribuida.

La figura 5 muestra un servidor 120 de sitio en el que coexisten múltiples proveedores diferentes para una variedad de fuentes de información. La organización estructural del servidor 120 de sitio facilita la capacidad de una red de información distribuida de la invención para acceder y extraer información útil desde una fuente de información particular una vez encontrada. Con referencia a la figura 5, el servidor 120 de sitio tiene un gestor 122 de proveedores que encamina una consulta de búsqueda entrante a uno apropiado o unos apropiados de los cinco proveedores mostrados en el ejemplo presentado. Los proveedores incluyen un proveedor 124 a una base 126 de datos A de comercio electrónico y a una base 128 de datos A B2B, un proveedor 130 a un sistema 132 de archivos de WINDOWS, un proveedor 134 a un sistema 136 de archivos UNIX, un proveedor 138 a una base 140 de datos de contenido, y un proveedor 142 a una base 144 de datos B de comercio electrónico. Cada uno de los proveedores 124, 130, 134, 138 y 142 tiene un elemento 124a, 130a, 134a, 138a, y 142a de acceso respectivo. Un elemento de acceso puede encontrar, abrir, escribir y leer un objeto con independencia del tipo de plataforma o almacén de datos. (Un almacén de datos es un mecanismo de almacenamiento, tal como un sistema de archivos, base de datos, sistema de correo electrónico o archivo zip, que puede contener datos en un formato organizado). Un elemento de acceso también tiene la capacidad de "rastrear" (es decir, examinar el contenido de) un almacén de datos o buscar un único objeto de datos. (Un objeto de datos es un único archivo, un objeto en una base de datos, un mensaje de correo electrónico, un resultado de búsqueda o un fragmento de memoria). Los proveedores apropiados para responder a una consulta de búsqueda particular usan sus elementos de acceso para consultar sus fuentes de información o almacenes de datos asociados. Los elementos de acceso traducen entre el lenguaje de consulta de un servidor raíz de la red de información distribuida y el lenguaje de consulta de un almacén de datos. Esta implementación facilita el acceso a cualquier fuente de información y se describe detalladamente a continuación con referencia a las figuras 6 y 7.

Los elementos 130a y 134a de acceso de sistema de archivos usan un gestor 146 de analizadores, que funciona como un intérprete de lenguaje informático y en el ejemplo presentado incluye seis analizadores equipados para reconocer documentos en seis formatos de archivos de software diferentes. Un analizador sabe cómo leer el contenido de un objeto de datos y después extraer metadatos y almacenarlos en un formato común. Los seis analizadores incluyen analizadores de documentos de WORD, documentos de EXCEL, imágenes JPG, audio MP3, POWERPOINT y PDF. Con independencia de dónde y cómo está almacenado un archivo particular, el gestor 146 de analizadores dirige el archivo al analizador apropiado. Por ejemplo, si un archivo representa un documento WORD, el analizador de documentos de WORD extrae los metadatos para el proveedor. Los proveedores, junto con el gestor 146 de analizadores, habilitan el acceso a cualquier tipo de información incluyendo: páginas web estáticas, documentos de procesador de textos u hoja de cálculo, imágenes, música, vídeo, e información de base de datos legada. Los proveedores son ampliables para tratar automáticamente nuevos tipos de datos.

Los proveedores de la red de información distribuida permiten que la propia fuente de información conserve la propiedad de todos los datos. Los proveedores actúan como una ventana, directamente al interior de la fuente de datos, habilitando de este modo a las fuentes de información para que controlen quién tiene acceso a información particular y para que controlen cómo se muestran los resultados.

El papel de un elemento de acceso proviene de la existencia de datos en numerosas formas y en numerosas ubicaciones en numerosas plataformas. Según se expone anteriormente, la presente realización implementa una técnica que accede y analiza los datos de manera consistente y segura y después almacena los metadatos en un formato común. Las figuras 6 y 7 respaldan la siguiente explicación de esta técnica. La figura 6 es un diagrama de bloques de un gestor de analizadores de servidores de sitio ejemplar y sus analizadores para un elemento de acceso a archivos y su almacén de datos. La figura 7 es un diagrama de bloques que muestra más detalladamente la estructura y organización de un gestor de proveedores con siete elementos de acceso y un gestor de analizadores con siete analizadores.

Con referencia a la figura 6, un servidor 200 de sitio funciona para entregar, a un gestor 202 de analizadores, información desde un almacén 204 de datos, a través de un elemento 206a de acceso. (El elemento 206a de acceso es uno de múltiples elementos de acceso mostrados en la figura 7). Un proveedor (no mostrado) en el servidor 200 de sitio también está conectado a la base 208 de datos en una disposición estructural análoga a la mostrada para el servidor 120 de sitio y las bases 126, 128, 140 y 144 de datos en la figura 5. El gestor 202 de analizadores dirige información a múltiples analizadores, incluyendo, por ejemplo, un analizador 210 de documentos de WORD; un analizador 212 de correo electrónico; un analizador 214 de datos de base de datos; y analizadores 216 de otra información que representan conjuntamente a partir de la figura 7 un analizador 218 de página web, un analizador 220 de datos archivados, un analizador 222 de bases de datos LOTUS Notes o EXCHANGE, y un analizador 224 de imágenes, películas o música. Con referencia a la figura 7, un gestor 230 de elementos de acceso mantiene una lista de elementos de acceso registrados, de los que hay siete mostrados a modo de ejemplo. Los elementos 206a, 232a, 234a, 236a, 238a, 240a y 242a de acceso están asociados con, respectivamente, un almacén 206 de datos de sistema archivos, un almacén 232 de datos de sistema de correo electrónico, un almacén 234 de datos de archivos de red, un almacén 236 de datos de bases de datos, un almacén 238 de datos de LOTUS Notes, un almacén 230 de datos de servidor de Internet y un almacén 232 de datos de archivos zip.

Con referencia a las figuras 6 y 7, la técnica para acceder y analizar datos es un mecanismo para recorrer (es decir, leer un sistema de archivos) un almacén de datos y analizarlo, con independencia de la ubicación de los datos o su tipo. Al tratar almacenes de datos y objetos de datos de manera genérica, el sistema distribuye un objeto genérico que

representa un objeto de datos. Este objeto de datos puede accederse a sí mismo desde el almacén de datos cargando y guardando la información y analizar sus datos en busca de propiedades ampliadas. El bloque 250 de proceso representa un evento de *spider* que inicia el proceso de acceder a un almacén de datos y analizarlo. Un evento de *spider* empieza con una ubicación de inicio y un elemento de acceso de inicio. Hay un elemento de acceso asociado con cada almacén de datos. Un elemento de acceso tiene la capacidad de rastrear un almacén de datos o buscar un único objeto de datos.

Un elemento de acceso recorre una lista de objetos en su almacén de datos y o bien crea un alias (denominado un "Moniker") a partir del objeto o bien carga otro elemento de acceso para procesar el objeto. Un Moniker es un objeto que envuelve un objeto de datos, que puede ser un archivo, un fragmento de datos en memoria o un enlace abstracto a cualquier tipo de objeto. El Moniker es lo que se pasa entre los elementos de acceso, analizadores, servidores y clientes. Los elementos de acceso tienen una interfaz encontrar primero/encontrar siguiente que devuelve Monikers o referencias a otros elementos de acceso. Los elementos de acceso también tienen una interfaz de usuario con la capacidad de incluir o excluir datos y establecer ubicaciones de inicio y fin al procesar una fuente de datos.

El gestor 230 de elementos de acceso mantiene una lista de todos los elementos de acceso registrados y los carga según sea necesario. El Moniker se crea por el elemento de acceso. El elemento de acceso carga entonces indirectamente un analizador. El Moniker puede compartirse entre servidores o clientes remotos. Con un Moniker, puede pedirse información de archivo, propiedades ampliadas o cualquier otra información dinámica.

El gestor 202 de analizadores puede cargar un analizador para un tipo de archivo dado. Un analizador procesa un archivo extrayendo datos. Un analizador puede soportar numerosos tipos de datos o un único tipo de datos específico. Puede haber múltiples analizadores que soportan el mismo tipo de datos, y el gestor 202 de analizadores determina el mejor analizador basándose en la plataforma, los componentes instalados, u otros factores. Cualquier analizador puede usar cualquier elemento de acceso.

El uso de un elemento de acceso, un analizador y un Moniker proporciona una capacidad de recorrer cualquier almacén de datos o almacenes de datos integrados en otros almacenes de datos (por ejemplo, archivos zip en sistemas de archivos o correo electrónico) y abrir y analizar datos con independencia del formato del archivo.

La figura 8 es un diagrama de bloques que muestra una red 300 de información distribuida compuesta por varias redes de aplicaciones, que expone una arquitectura de Internet distribuida que representa un híbrido de modelos centralizados y entre pares. Con referencia a la figura 8, la red 300 de información distribuida incluye una red 302 interna compuesta por un servidor 304 raíz, un servidor 306 intermedio, un servidor 308 de sitio de compras alojado de comercio electrónico, servidores 310 de sitio de alimentación de datos de comercio electrónico y servidores 312, 314 y 316 subraíz públicos de información. El servidor 304 raíz opera de la manera descrita anteriormente para el servidor 14 raíz de la figura 1, y el servidor 306 intermedio mejora los metadatos recopilados a partir de diversos servidores en la red 300.

En particular, el servidor 306 intermedio usa modelos, atributos de modelo, y conjuntos de campos para realizar diversas manipulaciones, comparaciones, ordenaciones de información y otros procesos, para presentar al usuario cliente la información recuperada de una manera que salva el hueco de información inherente a los motores de búsqueda actuales de la técnica anterior. Como se indica en la figura 8, para administrar su operación, el servidor 306 intermedio está organizado por clientes, tal como comercio electrónico, empresa a empresa (B2B) e información de comunidad. El comercio electrónico B2B se refiere a operaciones comerciales que se realizan entre una empresa y su cadena de suministro o entre una empresa y otros clientes finales de la empresa. El servidor 310 de sitio de compras alojado de comercio electrónico es un mercado en línea que presenta productos directamente a los consumidores. El servidor 310 de sitio proporciona, a través del servidor 304 raíz, acceso directo en tiempo real a cada catálogo de comerciante suscrito que conduce a un listado de productos real, en vez de a un enlace a un sitio web. La tecnología de proveedor de información descrita anteriormente habilita una adaptación a medida del cliente avanzada de la información tal como establecimiento de precios dinámico y filtrado por categorías. Los servidores 310 de sitio de alimentación de datos de comercio electrónico almacenan, en la red 302 interna, información proporcionada por el cliente como una adaptación a proveedores de información que no quieren búsquedas en directo realizadas en sus sitios.

Los servidores 312, 314 y 316 subraíz públicos de información representan tres ejemplos de servidores subraíz para grupos de interés de comunidad pública, cada uno de los cuales tiene potencialmente un número creciente de proveedores de información y consumidores de información. Estos servidores subraíz, que están alojados y administrados por un gestor de red y operan en cooperación con el servidor 304 raíz, dan acceso directo en tiempo real a cada fuente de información en su red, para garantizar que toda la información actual es accesible sin que se devuelvan enlaces rotos.

El sitio 308 de compras alojado de comercio electrónico y los servidores 312, 314, 316 y 354 subraíz de comunidad de información representan un portal de información que abre Internet de manera que cualquier usuario puede publicar cualquier tipo de información o acceder a cualquier tipo de dispositivo. El portal de información puede soportar un número indefinido de tipos de información (por ejemplo, sitios web, servidores de archivos, bases de datos y archivos de imágenes) y cualquier número de fuentes de información, con independencia de si están estructuradas o no estructuradas.

El servidor 304 raíz tiene múltiples servidores de nivel uno, incluyendo un servidor 318 A de sitio de comercio y un servidor 320 B de sitio de comercio.

El servidor 318 A de sitio de comercio representa un servidor de nivel uno de comercio electrónico B2B con un proveedor 322 de comercio electrónico y un proveedor 324 B2B que son análogos a los proveedores descritos con referencia al servidor 16 de sitio de la figura 3. El servidor 318 A de sitio de comercio tiene un servidor 326 A1 de nodo de sitio hijo de comercio de nivel dos, que tiene un enlace de comunicación con el proveedor 322 de comercio electrónico y representa una red de información privada de comercio electrónico. El servidor 326 A1 de nodo de sitio hijo de comercio tiene un proveedor 328 de comercio electrónico y un proveedor 330 de información que son análogos a los proveedores descritos con referencia al servidor 22 de nodo de sitio hijo de la figura 4. El servidor 326 de nodo de sitio hijo de comercio es una red interna privada en la que, por ejemplo, los empleados de la compañía propietaria del servidor A de sitio de comercio pueden acceder a documentos propietarios internos en toda la empresa, tales como documentos EXCEL. El servidor 318 A de sitio de comercio se muestra con un enlace de comunicación con un cliente 332 de compras privado de comercio electrónico que compra sólo los productos de la entidad que posee el servidor A de sitio de comercio y sus sitios hijos.

El servidor 320 B de sitio de comercio representa un servidor de nivel uno de comercio electrónico del lado de suministro B2B y de comercio electrónico B2B con un proveedor 334 de comercio electrónico y un proveedor 336 B2B que son análogos a los proveedores descritos con referencia al servidor 16 de sitio de la figura 3. El servidor 320 B de sitio de comercio tiene dos servidores 338 y 340 de nodo de sitio hijo de nivel dos, ambos de los cuales tienen enlaces de comunicación con el proveedor 236 B2B y representan suministradores B2B. Los dos servidores 338 y 340 de suministrador B2B pueden establecer una conexión del lado de suministro B2B, mediante la cual la entidad propietaria del servidor 320 B de sitio de comercio puede comprar suministros. El servidor 320 B de sitio de comercio se muestra con un enlace de comunicación con un cliente 342 de compras privado B2B que compra sólo los productos de la entidad que posee el servidor 320 B de sitio y sus sitios hijos.

Un cliente 350 de compras de comercio electrónico y un cliente 352 de compras de portal B2B compran cada uno en múltiples mercados a través del servidor 304 raíz. El cliente 350 de compras de comercio electrónico habilita las compras al por menor de empresa a consumidor (B2C) de múltiples sitios en múltiples mercados. El cliente 352 de compras de portal B2B habilita compras B2B de múltiples sitios en un mercado dado y de este modo crea una oportunidad de mercado para una red de participantes comerciantes ilimitada para crear un catálogo de productos en red en directo y dinámico.

La figura 8 muestra servidores 312, 314 y 316 subraíz públicos de información y un servidor 354 subraíz privado de información, asociado con lo que se denominan grupos ovoides de aplicación de información, cada uno de los cuales está compuesto por un cliente y un servidor de nodo. Un grupo 356 ovoide de aplicación de información tiene un enlace de comunicación con el servidor 312 subraíz público de información; un grupo 358 ovoide de aplicación de información tiene un enlace de comunicación con los servidores 356 y 358 subraíz públicos de información; y un grupo 360 ovoide de aplicación de información está asociado con el servidor 354 subraíz privado. Se establecen enlaces 362, 364 y 366 de comunicación entre pares (P2P), respectivamente, entre los grupos 356 y 358 ovoides de aplicación de información, entre los grupos 358 y 360 ovoides de aplicación de información, y entre el grupo 356 ovoide de aplicación de información y el proveedor 330 de información del servidor 326 A1 de sitio hijo de comercio. Los enlaces de comunicación P2P son conexiones entre ordenadores autónomos mediante los que puede descargarse un archivo desde uno de los ordenadores al otro sin la actuación de un servidor raíz. El servidor 354 subraíz privado de información aloja y administra su propio servidor y determina quién obtiene acceso, derechos y privilegios asociados con el mismo.

La figura 9 es un diagrama de bloques que muestra detalladamente los componentes y estructura de un grupo ovoide de aplicación de información en asociación operativa con el servidor 304 raíz de la red 302 interna. Con referencia a la figura 9, un servidor de registro-servidor raíz representa el papel desempeñado por el servidor 304 raíz; la comunidad 1 subraíz y la comunidad 2 subraíz representan los papeles desempeñados por dos cualesquiera de los servidores 312, 314 y 316 subraíz públicos de información; y la comunidad 3 subraíz representa el papel desempeñado por el servidor 354 subraíz privado de información. Un grupo ovoide de aplicación de información está compuesto por dos partes, que se indican mediante la línea horizontal que divide en dos partes cada uno de los grupos 356, 358 y 360 ovoides de aplicación de información en la figura 8. La parte de cliente de un grupo 400 ovoide de aplicación de información ejemplar incluye cómo sus componentes un ordenador 402 de usuario cliente, tal como un PC y un perfil 404 de usuarios locales en un sistema 406 de archivos. La capacidad de compartir archivos es un derecho del usuario, y el perfil 404 registra las identificaciones de los usuarios locales autorizados por el usuario cliente. El sistema 406 de archivos almacena archivos descargados desde los servidores de comunidad objetivo. La parte de servidor del grupo 400 ovoide de aplicación de información incluye como sus componentes el servidor 200 de sitio; el gestor 2020 de analizadores y sus analizadores 210, 212, 214 y 216 asociados; el almacén 204 de datos y su elemento 206 de acceso asociado; y una base 208 de datos. Esta configuración de componentes de servidor es la misma que la presentada en la figura 6; por tanto, por motivos de claridad, se usan los mismos números de referencia para indicar componentes comunes en las figuras 6 y 9. En una realización preferida, las funciones de las partes de cliente y servidor se combinan de manera que residen en la misma plataforma.

Según la realización, para el grupo 400 ovoide de aplicación de información, una búsqueda por un usuario cliente hace que una consulta de búsqueda llegue al servidor 200 de sitio de comunidad, que está incluido en el proceso de búsqueda y produce un archivo a partir del almacén 204 de datos para la entrega al usuario cliente.

Un asunto problemático surge en una red P2P, tal como la establecida por cualquiera de los enlaces 362, 364 y 366 de comunicación P2P, proviene del hecho de que el contenido puede residir en cualquier servidor del mismo nivel en la red P2P. Estos servidores carecen de conocimiento específico de otros servidores del mismo nivel en la red, aparte de un servidor de referencia que funciona como la fuente autorizadora de la información en red (es decir, un servicio de directorio). Para evitar que clientes del mismo nivel no autorizados busquen en servidores del mismo nivel en la red P2P, la realización implementa un procedimiento que indica a un servidor del mismo nivel que un cliente del mismo nivel que solicita una búsqueda está autorizado para hacerlo.

El procedimiento se lleva a cabo mediante la operación del servidor 304 de registro-servidor raíz de la figura 9, que es un servidor central conocido por todos los clientes y usado como un repositorio para claves públicas dentro de la red P2P. Al unirse a la red P2P por primera vez, un cliente pasa al servidor 304 de registro-servidor raíz una parte de clave pública del par de claves pública/privada generado por el cliente, junto con una dirección de correo electrónico y otra información según se requiera por un administrador de red. El cliente se identifica como uno de los grupos ovoides de aplicación de información en las figuras 8 y 9. El cliente en ese momento obtiene la clave pública que identifica al servidor 304 de registro-servidor raíz y almacena su clave pública para futura referencia. El proceso de conexión de registro se indica mediante la línea discontinua con punta de flecha entre el servidor de comunidad 1 subraíz y el servidor 200 de sitio y la línea continua que conecta el servidor de comunidad 1 subraíz y el servidor 304 de registro-servidor raíz en la figura 9.

La figura 10 es un diagrama de flujo del proceso de seguridad y autenticación de sesión llevado a cabo en una red P2P. Cada uno de los servidores de comunidad 1 a 3 subraíz de la figura 9 replica las funciones de autorización del servidor 304 de registro-servidor raíz. Por tanto, estos servidores de comunidad almacenan las claves públicas de los usuarios cliente de la red P2P. Con referencia a la figura 10, la siguiente vez tras el registro, el cliente establece comunicación con el servidor de comunidad 1 subraíz para solicitar una cadena de bits de desafío. El servidor de comunidad 1 subraíz genera en respuesta una cadena de bits aleatorios y la envía al cliente como una cadena de bits de desafío. El cliente cifra entonces la cadena de bits de desafío usando la clave privada del cliente y devuelve la cadena de bits de desafío cifrada al servidor de comunidad 1 subraíz. El servidor de comunidad 1 subraíz descifra entonces la cadena de bits de desafío devuelta por el cliente usando la clave pública que el servidor de comunidad 1 subraíz tiene en archivo para el cliente y compara los resultados del descifrado con la cadena de bits de desafío original. Para una verificación satisfactoria, el resultado de descifrado de la cadena de bits de desafío con la clave pública coincide con la cadena de bits de desafío original, proporcionando de este modo la identidad del cliente.

Una vez que se ha establecido la identidad del cliente, el servidor de comunidad 1 subraíz devuelve al cliente un testigo de acceso que permite al cliente consultar otros servidores del mismo nivel en la red P2P. Este testigo de acceso incluye, por ejemplo, la dirección IP de la que informó el cliente durante el desafío/respuesta y una marca de fecha y hora desde el servidor de comunidad 1 subraíz. El testigo de acceso se firma entonces, usando la clave privada del servidor de comunidad 1 subraíz.

Cuando desea buscar información en un servidor del mismo nivel objetivo, el cliente pasa el testigo de acceso junto con el paquete de petición de consulta. El servidor 200 del mismo nivel objetivo que recibe la petición valida entonces el testigo de acceso. El proceso de validación puede adoptar una de dos formas. Puesto que conoce la clave pública del servidor de comunidad 1 subraíz, el servidor 200 del mismo nivel objetivo puede por sí mismo validar el testigo de acceso. Alternativamente, puede pasarse el testigo de acceso al servidor de comunidad 1 subraíz, y validarse en el mismo. Si la marca de fecha y hora se usa para crear un testigo de acceso con una duración limitada, la comprobación de vuelta en el servidor de comunidad 1 subraíz eliminaría cualquier problema en cuanto a husos horarios. Una determinación de un testigo de acceso válido da como resultado la entrega de una petición de descargar datos, acompañada por el testigo de acceso, al servidor 200 del mismo nivel objetivo que, en respuesta, descarga los datos al cliente 402.

La prueba de la identidad de cliente se emprende al inicio de cualquier sesión con un sistema remoto, de modo que si se realiza una búsqueda durante una sesión que es diferente de una sesión de transferencia de archivos, volvería a enviarse el testigo de acceso, y volvería a verificarse cuando se inicie la sesión de transferencia de archivos.

Para exhibir la capacidad adicional de la red 300 de información distribuida, la figura 9 muestra con una línea discontinua con punta de flecha una conexión de consulta de comunidad entre el cliente 402 y el servidor de comunidad 3 subraíz privado, para ilustrar la capacidad del cliente 402 para buscar en un servidor de comunidad privado. Un proceso de autenticación se emprende para abrir una sesión con un servidor de comunidad privado.

Otro asunto problemático surge en conexión con un entorno distribuido en el que se comparten archivos u otra información. Puesto que los permisos de recurso compartido residen preferiblemente en la fuente de datos, los riesgos de seguridad provienen de un atacante potencial que desee compartir contenido no aprobado y que tenga acceso físico al ordenador que contiene los datos y la información de recurso compartido. Esta situación permite dos clases de ataque. La primera clase es la sustitución de la propia fuente de datos. Esto se consigue con gran facilidad sobrescribiendo un archivo compartido con un archivo no aprobado. La segunda clase de ataque es la modificación de la información de recurso compartido, que normalmente residirá en una base de datos. Alterar estos datos puede permitir que los datos apunten a un archivo no aprobado en vez de al contenido aprobado.

La figura 11 es un diagrama de flujo que define en términos generales las cinco etapas de un proceso para proporcionar seguridad para compartir archivos en una red P2P. Con referencia a la figura 11, el servidor de comunidad 1 subraíz que funciona como un administrador tiene, como se describe con referencia a la figura 10, autoridad de aprobación para contenido y se identifica mediante un par de claves pública/privada. La parte de clave pública de este par de claves se distribuye a todos los servidores de nodo del mismo nivel en la red P2P.

La etapa 1 del proceso representa un evento cuando un usuario desea compartir contenido. La información acerca de tal contenido (mostrada como información de fila 1 de la tabla de archivos del servidor de recursos compartidos) incluyendo el nombre del archivo, el tamaño del archivo y el *hash* del archivo se envía al servidor de comunidad 1 subraíz (de autorización). (Un "*hash*" se forma mediante un algoritmo criptográfico, es una representación condensada del contenido de un archivo). El servidor de comunidad 1 subraíz examina el archivo para garantizar que el contenido es apropiado.

La etapa 2 conlleva el uso por el servidor de comunidad 1 subraíz de la información de fila 1 para acceder al archivo de forma remota. La etapa 3 conlleva la aprobación del archivo por el servidor de comunidad 1 subraíz, que aplica la función *hash* al nombre del archivo, al tamaño del archivo y al *hash* del archivo. Cuando aprueba el archivo para que se comparta, el servidor de comunidad 1 subraíz, usando su clave privada, firma la información que se le envió. La etapa 4 representa que la firma, junto con el contenido compartido, se almacenan en la tabla de archivos en el servidor de recursos compartidos.

La etapa 5 representa cuando un servidor de recursos compartidos recibe una petición para descargar un archivo de información compartida a un servidor del mismo nivel. El servidor de recursos compartidos recupera, en respuesta, el nombre del archivo, obtiene el tamaño del archivo a partir del sistema de archivos y calcula el *hash* del archivo. Se aplica entonces la función *hash* a estos tres valores y se comparan frente al *hash* firmado descifrado devuelto desde el servidor de comunidad 1 subraíz. Si cualquiera de estos valores no coincide, el archivo no se pone a disposición del servidor del mismo nivel que solicita la descarga. En caso contrario, el archivo se pone a disposición del servidor del mismo nivel.

Aunque se describe con referencia a una red P2P, el proceso de seguridad para compartir archivos puede implementarse en cualquier red en la que un servidor puede conseguir un acceso controlado a un archivo que reside en un servidor ubicado de forma remota.

Resultará obvio para los expertos en la técnica que pueden realizarse numerosos cambios en los detalles de las realizaciones anteriormente descritas de esta invención, sin alejarse de los principios subyacentes de la misma. Como un primer ejemplo, pueden combinarse las funciones de un cliente (por ejemplo, *applet* cliente) y un servidor raíz, de modo que residan en la misma plataforma. Como un segundo ejemplo, un *applet*, una aplicación, un explorador de red, u otro tipo de cliente de sistema operativo, pueden usarse para iniciar una consulta o búsqueda de tema. El ámbito de la invención debe, por tanto, determinarse sólo por las siguientes reivindicaciones.

REIVINDICACIONES

1. Una red (10) de información distribuida construida para recopilar información a partir de sitios distribuidos a través de una red informática accesible de forma global, que comprende:

un servidor (14) raíz que almacena una lista de múltiples sitios (A, B, C) distribuidos, cada uno de los cuales está representado por metadatos correspondientes a contenido de información disponible directa o indirectamente;

múltiples sitios (A, B, C) distribuidos, cada uno de los cuales se implementa con un proveedor (16, 18, 20) de información que está ubicado de forma remota del servidor (14) raíz, almacenando el proveedor de información de cada uno de los sitios distribuidos, metadatos correspondientes al contenido de información que está disponible en un servidor (A, B, C) de o un servidor (A1, C1, C2, C2-1, C2-2) asociado con el proveedor de información, pudiendo recuperarse el contenido de información en respuesta a una petición de búsqueda de información perfilada desde el servidor raíz de resultados de búsqueda de una búsqueda realizada sobre el contenido de información; **caracterizada** por:

un enlace de comunicación de información perfilada entre el servidor (14) raíz y cada uno de los múltiples sitios (A, B, C) distribuidos, habilitando el enlace de comunicación de información perfilada la formación de una trayectoria para la entrega de los resultados de búsqueda del contenido de información al que corresponden los metadatos a un sitio de destino a partir de un sitio o sitios representados por los metadatos de la petición de búsqueda de información perfilada; y

un cliente (12) de sistema operativo que entrega al servidor (14) raíz una petición de búsqueda de información por un usuario, reuniendo el servidor raíz una lista que identifica a uno o más seleccionados de los servidores válidos para buscar contenido de información en los mismos, y recibe sin pasar a través del enlace de comunicación de información perfilada los resultados de búsqueda recuperados a partir del uno o más servidores seleccionados identificados en la lista en respuesta a la petición de búsqueda de información perfilada.

2. La red de información distribuida según la reivindicación 1, en la que los múltiples sitios (A, B, C) distribuidos están configurados para alojar y mantener su propio contenido de información mientras están disponibles para el acceso mediante peticiones de búsqueda de información que se originan desde fuentes de red informática accesible de forma global ubicadas de forma remota.

3. La red de información distribuida según la reivindicación 1, en la que el cliente (12) de sistema operativo comprende uno de un explorador de red, un *applet* o una aplicación.

4. La red de información distribuida según la reivindicación 1, en la que al menos uno de los múltiples sitios (A, B, C) distribuidos implementa una red de información local y comprende además:

un servidor (16, 18, 20) raíz local que almacena una lista de múltiples sitios locales distribuidos, cada uno de los cuales está representado por metadatos locales correspondientes a un contenido de información disponible directa o indirectamente; y

múltiples sitios (A1, C1, C2, C2-1, C2-2) locales distribuidos, cada uno de los cuales se implementa con un proveedor de información en el que se almacenan metadatos locales correspondientes al contenido de información que puede recuperarse en respuesta a una petición de búsqueda de información perfilada local de resultados de búsqueda de una búsqueda realizada sobre el contenido de información disponible en un servidor o servidores locales asociados con el proveedor de información.

5. La red de información distribuida según la reivindicación 1, en la que al menos uno de los múltiples sitios (A, B, C) distribuidos incluye múltiples niveles de servidores (16, 18, 20, 22, 24, 26, 28, 30) en los que buscar en respuesta a la petición de búsqueda de información perfilada.

6. La red de información distribuida según la reivindicación 1, en la que los múltiples sitios (A, B, C) distribuidos incluyen servidores de sitio para al menos uno de una red de comercio electrónico, una red de empresa a empresa, una red de sitio de suministro de empresa a empresa, una red entre pares o una red de información de comunidad.

7. La red de información distribuida según la reivindicación 1, en la que el servidor (14) raíz incluye una interfaz (52) de analizador de consulta que tiene un proveedor (54) de sitio y un proveedor (56) principal, identificando el proveedor (54) de sitio uno determinado o unos determinados de los múltiples sitios (A, B, C) distribuidos correspondientes a los metadatos que representan perfiles de tema indicativos del contenido de información e identificando el proveedor principal propiedades de temas representados por los perfiles de tema.

8. La red de información distribuida según la reivindicación 7, en la que los proveedores (54, 56) de sitio y principal están asociados de forma operativa con bases de datos de perfil de sitio y de tema respectivas y en la que la interfaz de

analizador de consulta accede y recupera el contenido de información a partir de las bases de datos de perfil de sitio y de tema para reunir un paquete de información que incluye la lista que identifica el uno o más servidores seleccionados válidos para buscar contenido de información en los mismos.

9. La red de información distribuida según la reivindicación 1, en la que al menos uno de los múltiples sitios (A, B, C) distribuidos implementa una red de información local entre pares y comprende además:

múltiples sitios locales del mismo nivel distribuidos, cada uno de los cuales se implementa con un proveedor de información en el que se almacenan metadatos locales del mismo nivel y pueden recuperarse en respuesta a una petición de búsqueda de información perfilada local del mismo nivel de resultados de búsqueda que pueden derivarse del contenido de información al que corresponden los metadatos locales del mismo nivel; y

un servidor (304) raíz local del mismo nivel, que almacena una lista de los múltiples sitios locales del mismo nivel distribuidos, cada uno de los cuales está representado por metadatos correspondientes al contenido de información al que corresponden los metadatos locales del mismo nivel, y emite un testigo de acceso en respuesta a la recepción de una petición de búsqueda por uno válido de los múltiples sitios locales distribuidos para proporcionar una trayectoria aprobada para la entrega de resultados de búsqueda local del mismo nivel que son sensibles a la petición de búsqueda.

10. La red de información distribuida según la reivindicación 9, en la que el testigo de acceso se emite según un proceso de cifrado y descifrado con un par de claves pública/privada.

11. La red de información distribuida según la reivindicación 1, en la que al menos uno de los múltiples sitios distribuidos implementa una red de información local y comprende además:

un servidor (304) raíz local que se identifica mediante un par de claves pública/privada de servidor raíz que tiene una parte de clave pública y que almacena una lista de múltiples sitios locales distribuidos, cada uno de los cuales está representado por metadatos locales correspondientes a contenido de información disponible directa o indirectamente y partes de clave pública de pares de claves pública/privada de sitio generadas por los múltiples sitios locales distribuidos; y

múltiples sitios locales distribuidos, cada uno de los cuales se implementa con un proveedor de información en el que se almacenan metadatos locales correspondientes a contenido de información que puede recuperarse en respuesta a una petición de búsqueda de información perfilada local de resultados de búsqueda que pueden derivarse del contenido de información al que corresponden los metadatos locales, y almacenando los múltiples sitios locales distribuidos la parte de clave pública de servidor raíz para su uso para llevar a cabo un algoritmo criptográfico para evitar cambios no autorizados en el contenido de información representado por los metadatos locales almacenados.

12. La red de información distribuida según la reivindicación 11, en la que los múltiples sitios locales distribuidos forman parte de una red entre pares.

13. La red de información distribuida según la reivindicación 1, que comprende además:

una interfaz de administración asociada con el servidor raíz; y

un enlace de comunicación de actualización de metadatos entre la interfaz de administración y los múltiples sitios distribuidos, para proporcionar a la interfaz de administración información de base de metadatos actualizada y de este modo habilitar al servidor raíz para alojar bases de metadatos actualizadas de tipos de información almacenada en la red de información distribuida.

14. Un procedimiento de recopilación de información a partir de múltiples sitios distribuidos a través de una red informática accesible de forma global, que comprende:

implementar, con cada uno de los múltiples sitios (A, B, C) distribuidos, un proveedor (16, 18, 20) de información que almacena metadatos correspondientes a contenido de información que está disponible en un servidor (A, B, C) de o un servidor (A1, C1, C2, C2-1, C2-2) asociado con el proveedor de información, pudiendo recuperarse el contenido de información en respuesta a una petición de búsqueda de información perfilada de resultados de búsqueda de una búsqueda realizada sobre el contenido de información;

establecer un enlace de comunicación de información perfilada entre un servidor (14) raíz que almacena una lista de múltiples sitios distribuidos, cada uno de los cuales está representado por metadatos correspondientes al contenido de información y cada uno de los múltiples sitios distribuidos, **caracterizado** por:

transmitir desde un cliente (12) de sistema operativo al servidor (14) raíz una petición de búsqueda de información perfilada de resultados de búsqueda de una búsqueda realizada sobre el contenido de información;

reunir el servidor raíz una lista que identifica a uno o más seleccionados de los servidores válidos para buscar contenido de información en los mismos; y

formar una trayectoria de comunicación para la entrega de los resultados de búsqueda a un sitio de destino desde el uno o más servidores seleccionados.

15. El procedimiento según la reivindicación 14, que comprende además proporcionar un enlace de comunicación con el cliente (12) de sistema operativo para entregarle los resultados de búsqueda recuperados a partir del sitio o sitios de destino en respuesta a la petición de búsqueda de información perfilada.

16. La red de información distribuida según la reivindicación 1, en la que cada uno de los múltiples sitios distribuidos incluye una interfaz de administración local y una base de metadatos local asociada, actualizando la interfaz de administración local su base de metadatos local asociada para proporcionar un registro actual del contenido disponible de forma local en el servidor.

17. La red de información distribuida según la reivindicación 16, en la que al menos uno de los múltiples sitios distribuidos incluye múltiples niveles de servidores y la base de metadatos local proporciona un registro actual que incluye el contenido disponible a partir de al menos un nivel de servidor de los múltiples niveles de servidores.

18. El procedimiento según la reivindicación 14, que comprende además:

implementar una interfaz de administración asociada con el servidor raíz; y

establecer un enlace de comunicación de actualización de metadatos entre la interfaz de administración y los múltiples sitios distribuidos, para proporcionar a la interfaz de administración información de base de metadatos actualizada y de este modo habilitar al servidor raíz para alojar bases de metadatos actualizadas de tipos de información almacenada en la red de información de distribución.

19. El procedimiento según la reivindicación 14, en el que cada uno de los múltiples sitios distribuidos incluye una interfaz de administración local y una base de metadatos local asociada, actualizando la interfaz de administración local su base de metadatos local asociada para proporcionar un registro actual del contenido disponible de forma local en el servidor.

20. El procedimiento según la reivindicación 19, en el que al menos uno de los múltiples sitios distribuidos incluye múltiples niveles de servidores y la base de metadatos local proporciona un registro actual que incluye el contenido disponible a partir de al menos un nivel de servidor de los múltiples niveles de servidores.

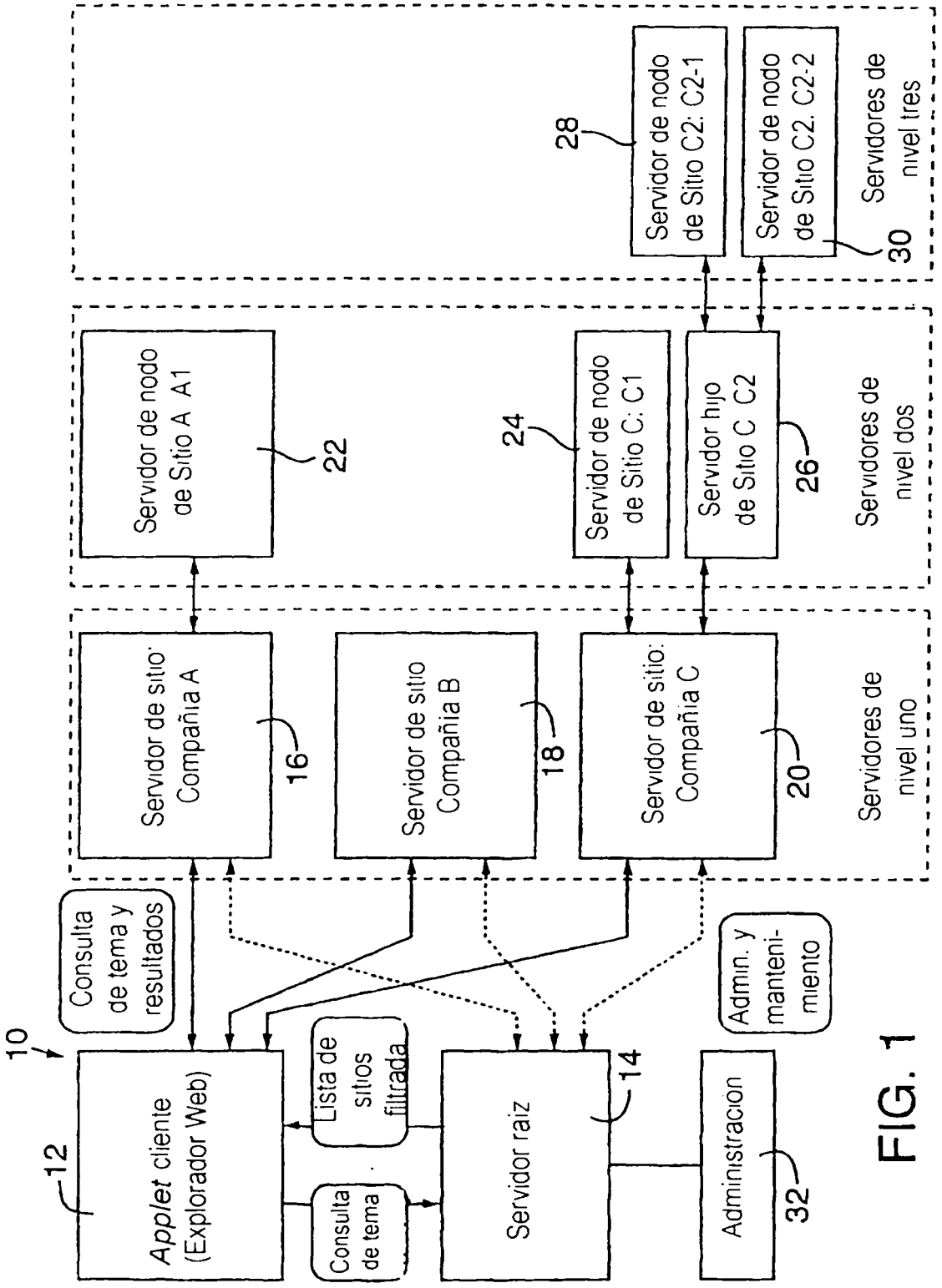

10
12
Applet cliente (Explorador Web)
Consulta de tema y resultados
Lista de sitios filtrada
Consulta de tema
Servidor raiz
14
Admin. y mantenimiento
Administracion
32
Servidor de sitio Compañia A
16
Servidor de sitio Compañia B
18
Servidor de sitio Compañia C
20
Servidores de nivel uno
Servidor de nodo de Sitio A A1
22
Servidor de nodo de Sitio C: C1
24
Servidor hijo de Sitio C C2
26
Servidores de nivel dos
28
Servidor de nodo de Sitio C2: C2-1
Servidor de nodo de Sitio C2. C2-2
30
Servidores de nivel tres

FIG. 1

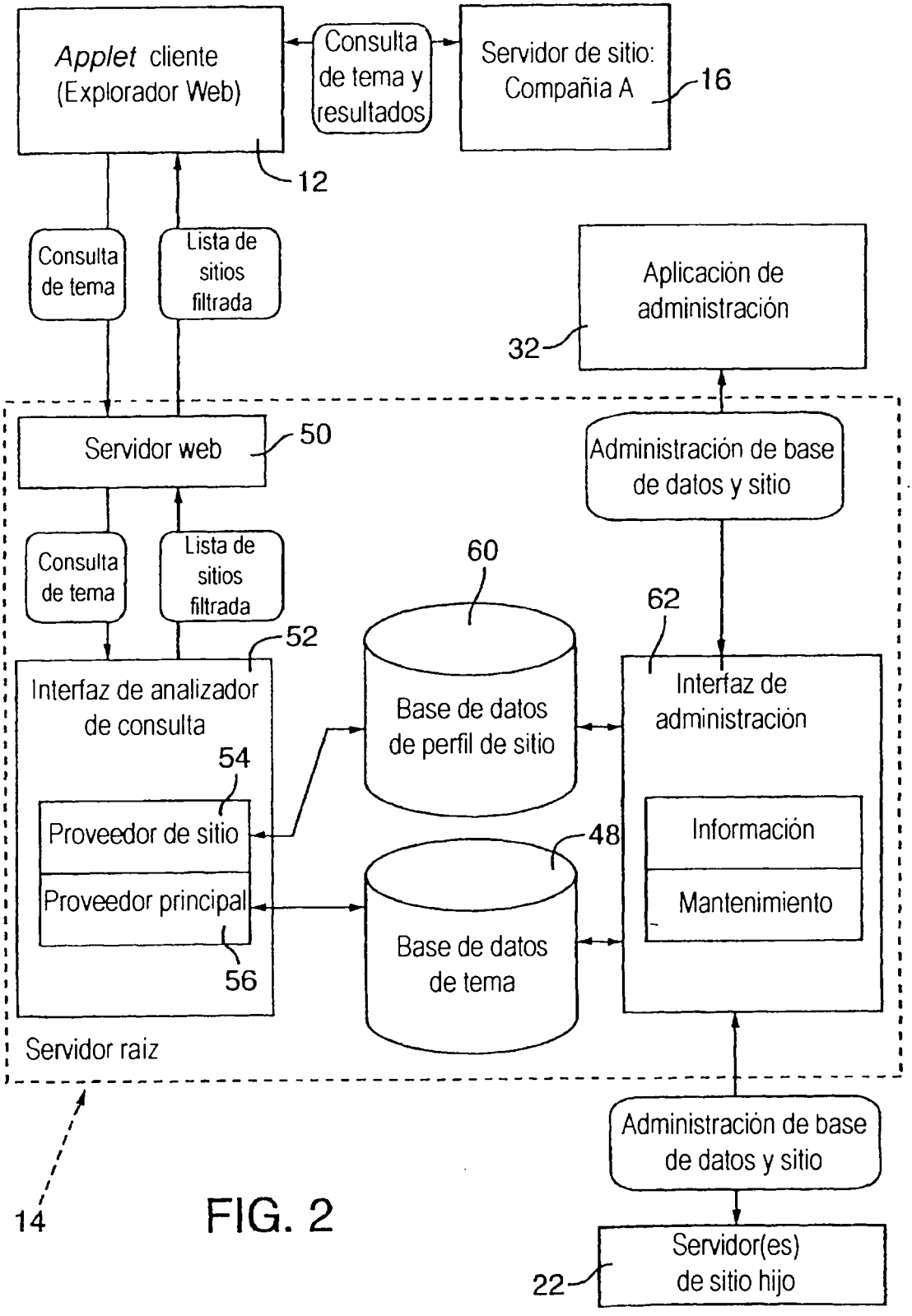
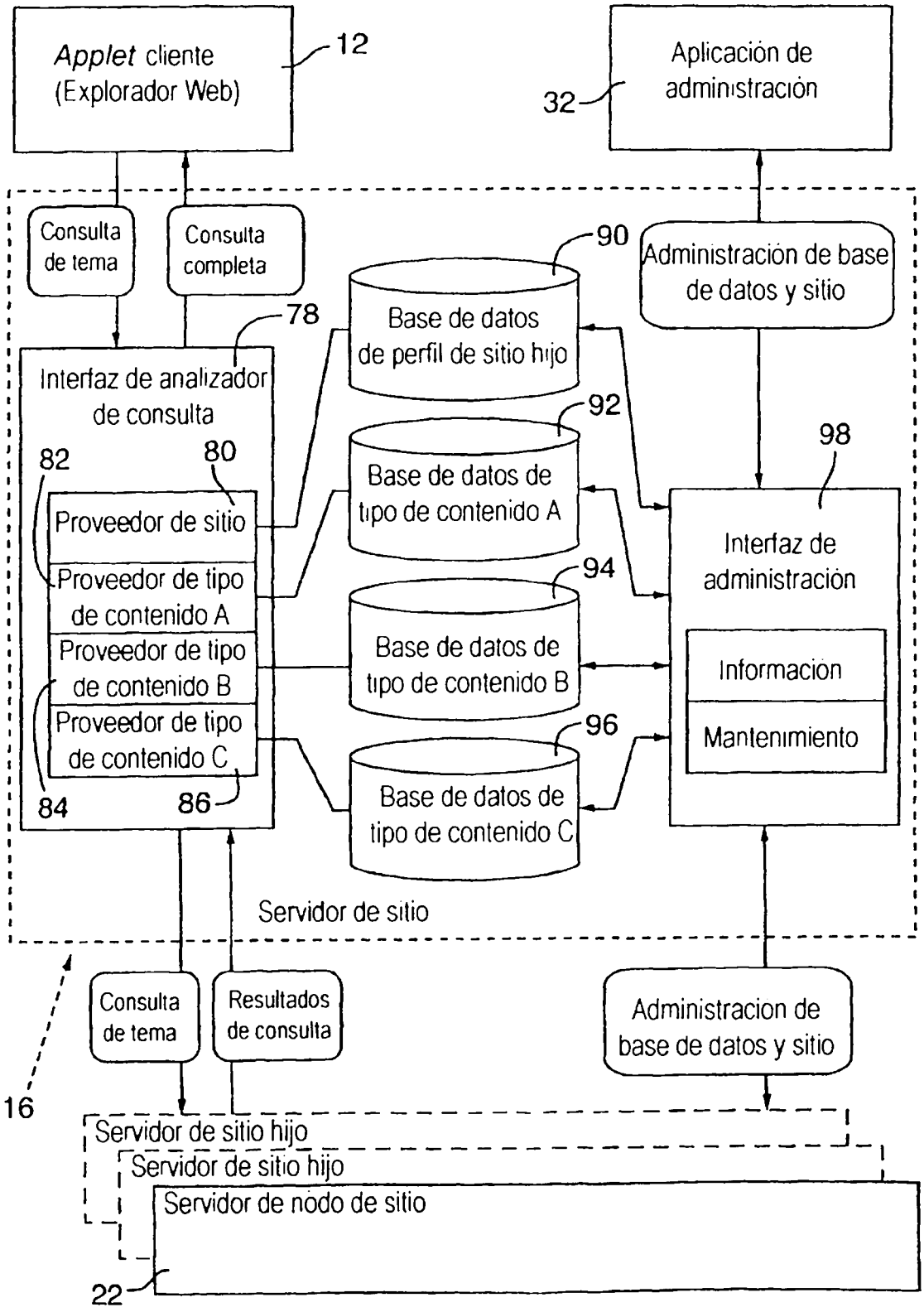

Applet cliente
(Explorador Web)
Consulta de tema y resultados
Servidor de sitio: Compañia A
16
12
Consulta de tema
Lista de sitios filtrada
Aplicación de administración
32
Servidor web
50
Administración de base de datos y sitio
Consulta de tema
Lista de sitios filtrada
60
62
52
Interfaz de analizador de consulta
Base de datos de perfil de sitio
Interfaz de administración
54
Proveedor de sitio
Información
48
Proveedor principal
Mantenimiento
Base de datos de tema
56
Servidor raiz
Administración de base de datos y sitio
14
Servidor(es) de sitio hijo
22

FIG. 2

FIG. 3


Applet cliente
(Explorador Web)
12
Aplicación de
administración
32
Consulta
de tema
Consulta
completa
78
Interfaz de analizador
de consulta
82
80
Proveedor de sitio
Proveedor de tipo
de contenido A
Proveedor de tipo
de contenido B
Proveedor de tipo
de contenido C
84
86
90
Base de datos
de perfil de sitio hijo
92
Base de datos de
tipo de contenido A
94
Base de datos de
tipo de contenido B
96
Base de datos de
tipo de contenido C
Administración de base
de datos y sitio
98
Interfaz de
administración
Información
Mantenimiento
Servidor de sitio
Consulta
de tema
Resultados
de consulta
Administracion de
base de datos y sitio
16
Servidor de sitio hijo
Servidor de sitio hijo
Servidor de nodo de sitio
22

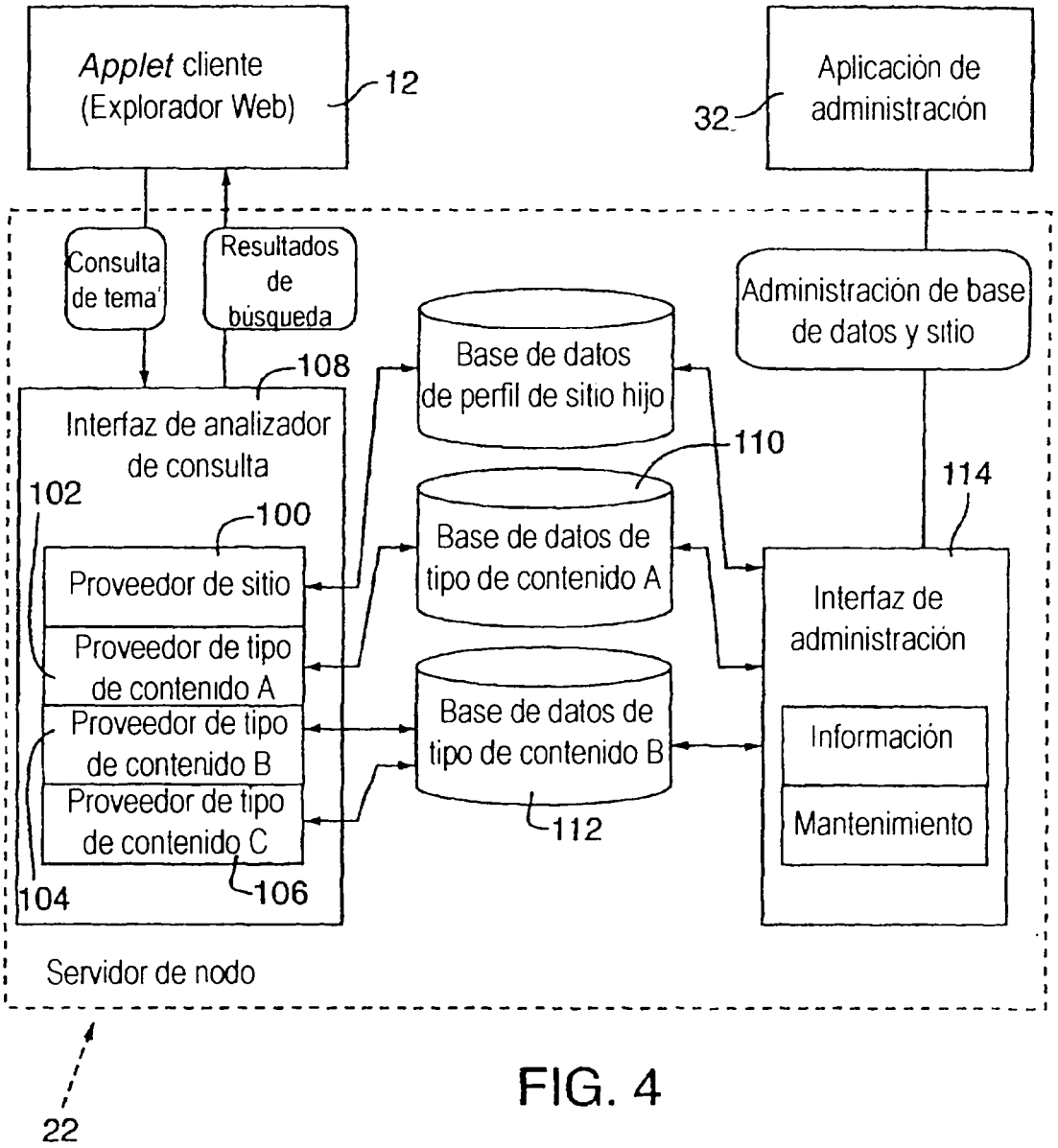
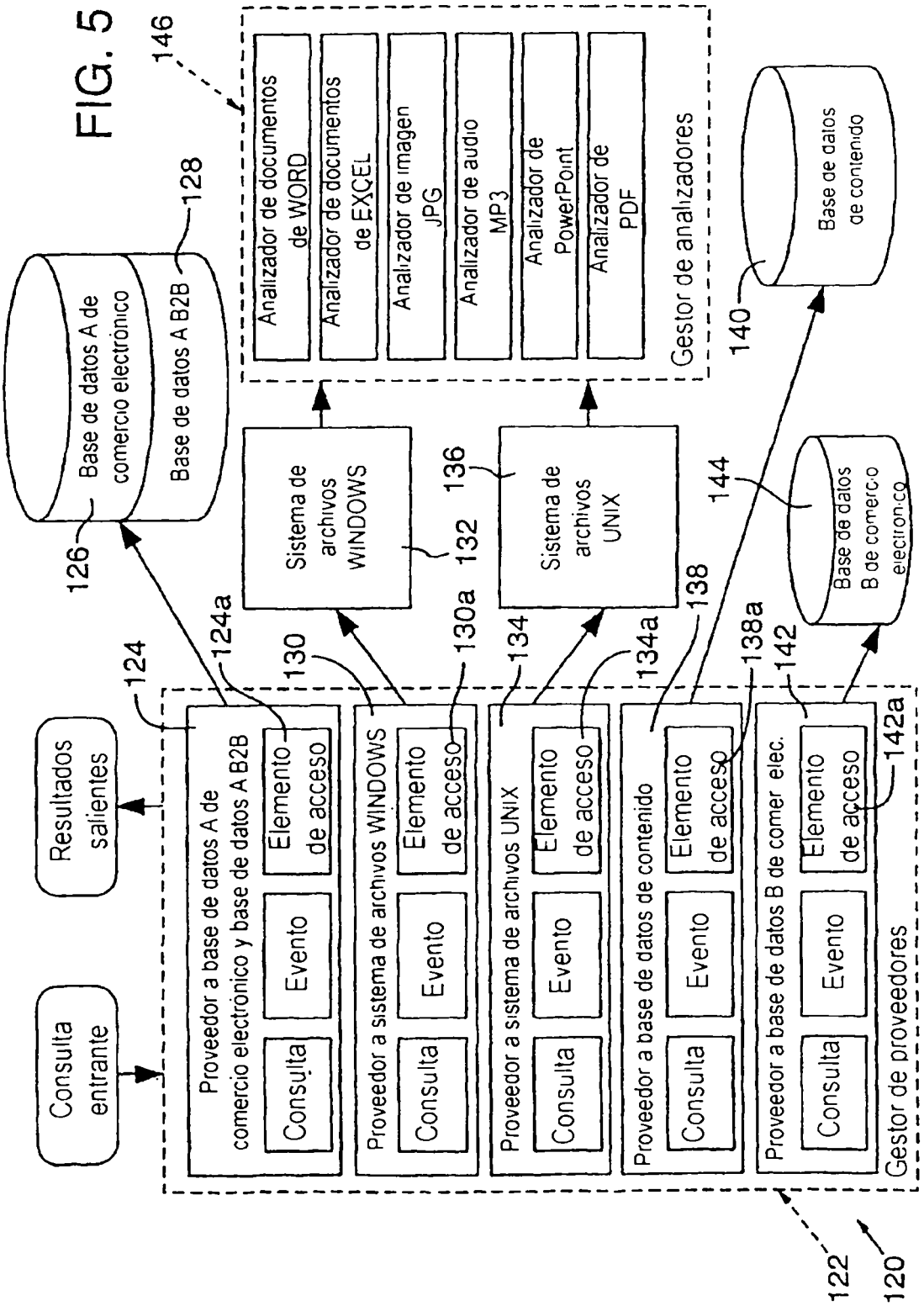

Applet cliente
(Explorador Web)
12
Aplicación de
administración
32
Consulta
de tema
Resultados
de
búsqueda
Administración de base
de datos y sitio
Base de datos
de perfil de sitio hijo
108
Interfaz de analizador
de consulta
110
114
102
100
Proveedor de sitio
Base de datos de
tipo de contenido A
Interfaz de
administración
Proveedor de tipo
de contenido A
Proveedor de tipo
de contenido B
Base de datos de
tipo de contenido B
Información
Proveedor de tipo
de contenido C
Mantenimiento
112
104
106
Servidor de nodo
22

FIG. 4

FIG. 5


Consulta entrante
Resultados salientes
Gestor de proveedores
Proveedor a base de datos A de comercio electrónico y base de datos A B2B
Consulta
Evento
Elemento de acceso
Proveedor a sistema de archivos WINDOWS
Proveedor a sistema de archivos UNIX
Proveedor a base de datos de contenido
Proveedor a base de datos B de comer elec.
Base de datos A de comercio electrónico
Base de datos A B2B
Sistema de archivos WINDOWS
Sistema de archivos UNIX
Gestor de analizadores
Analizador de documentos de WORD
Analizador de documentos de EXCEL
Analizador de imagen JPG
Analizador de audio MP3
Analizador de PowerPoint
Analizador de PDF
Base de datos de contenido
Base de datos B de comercio electrónico
120
122
124
124a
126
128
130
130a
132
134
134a
136
138
138a
140
142
142a
144
146

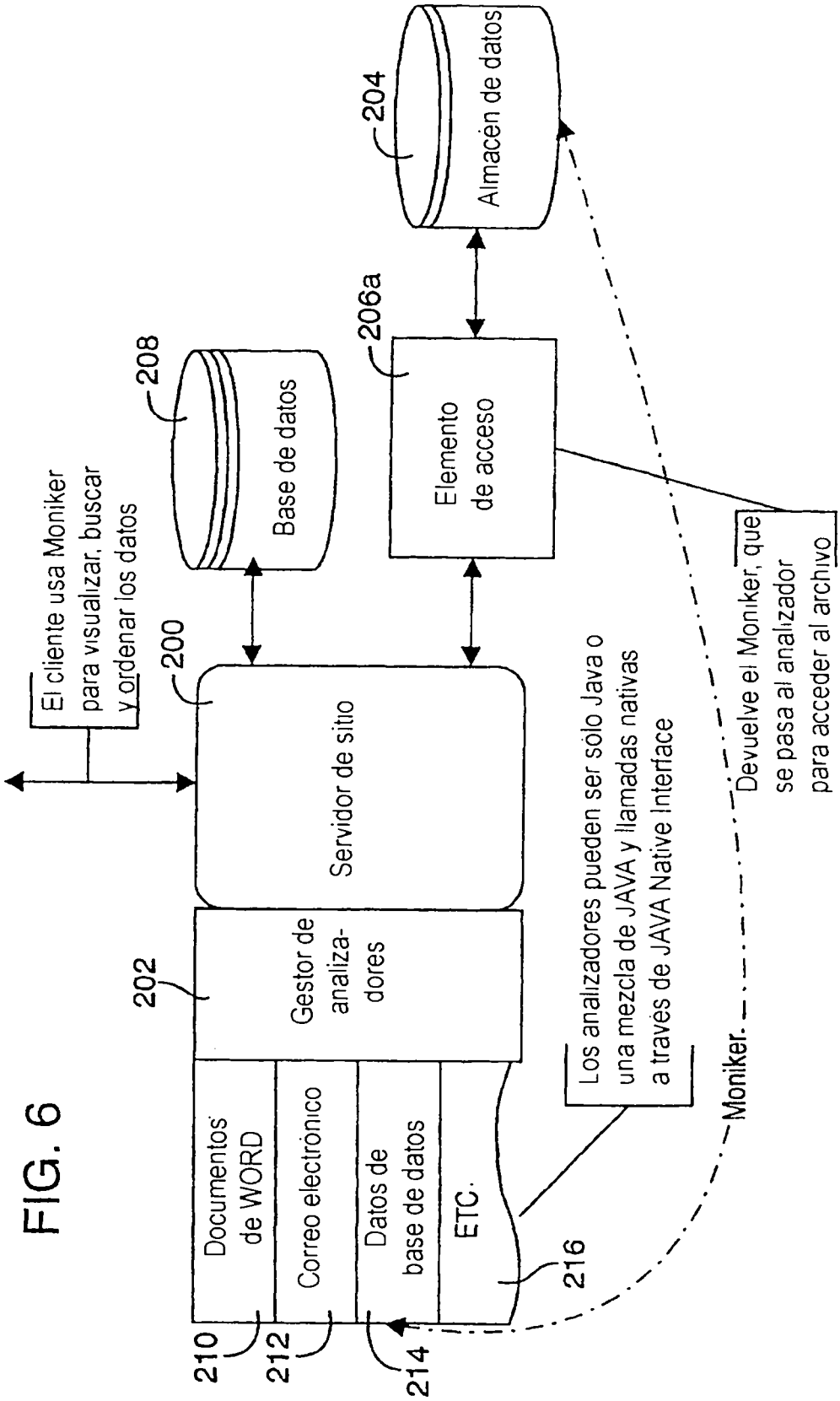
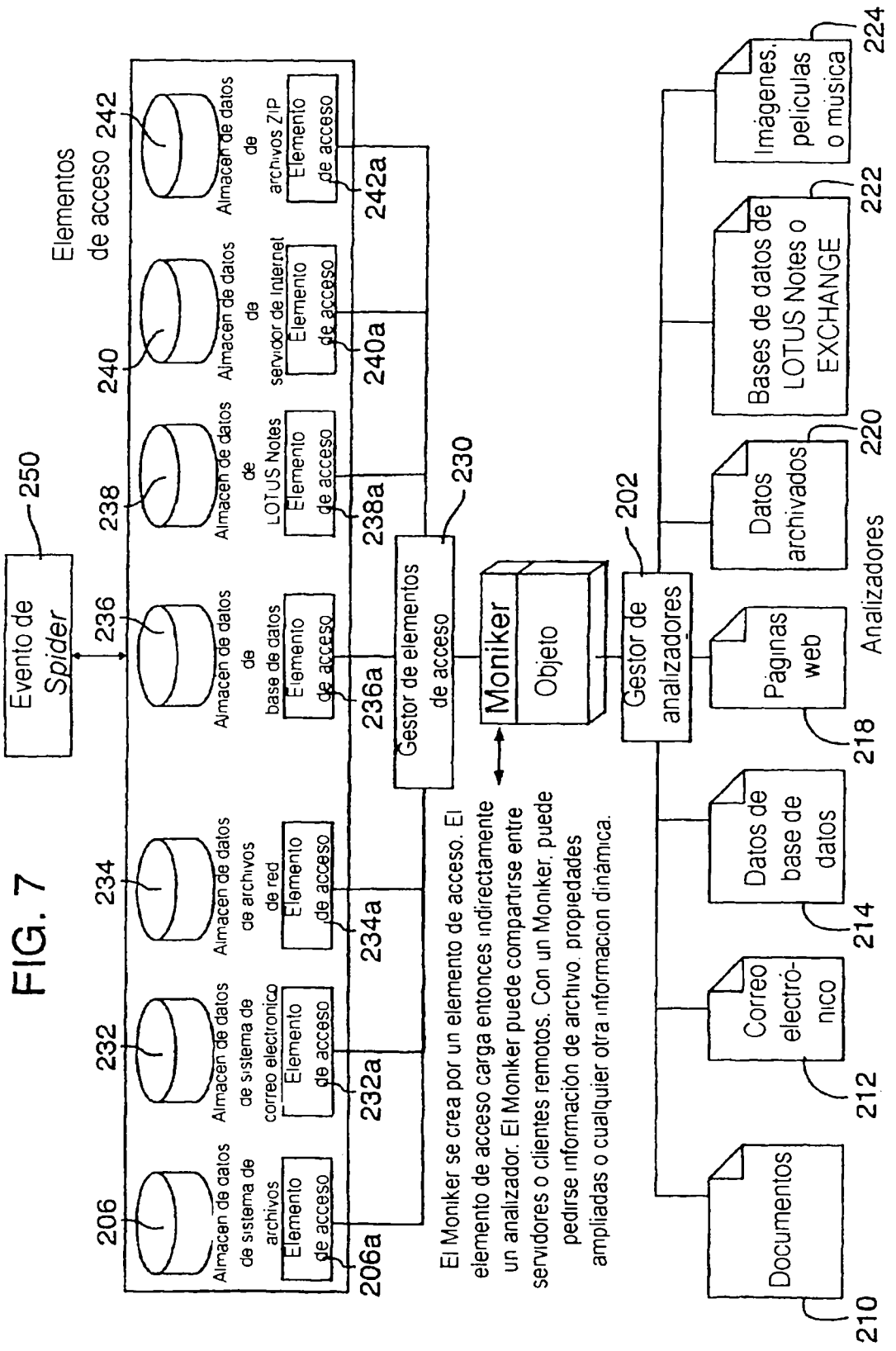

204
Almacén de datos
206a
Elemento de acceso
208
Base de datos
El cliente usa Moniker para visualizar, buscar y ordenar los datos
200
Servidor de sitio
Devuelve el Moniker, que se pasa al analizador para acceder al archivo
202
Gestor de analiza-dores
Los analizadores pueden ser sólo Java o una mezcla de JAVA y llamadas nativas a través de JAVA Native Interface
Moniker
Documentos de WORD
Correo electrónico
Datos de base de datos
ETC.
216
210
212
214

FIG. 6

FIG. 7


Evento de Spider
250
Elementos de acceso
242
240
238
236
234
232
206
Almacen de datos de sistema de archivos
Elemento de acceso
206a
Almacen de datos de sistema de correo electronico
232a
Almacen de datos de archivos de red
234a
Almacen de datos de base de datos
236a
Almacen de datos de LOTUS Notes
238a
Almacen de datos de servidor de Internet
240a
Almacen de datos de archivos ZIP
242a
Gestor de elementos de acceso
230
Moniker
Objeto
El Moniker se crea por un elemento de acceso. El elemento de acceso carga entonces indirectamente un analizador. El Moniker puede compartirse entre servidores o clientes remotos. Con un Moniker, puede pedirse información de archivo, propiedades ampliadas o cualquier otra información dinámica.
Gestor de analizadores
202
Analizadores
Documentos
210
Correo electrónico
212
Datos de base de datos
214
Páginas web
218
Datos archivados
220
Bases de datos de LOTUS Notes o EXCHANGE
222
Imágenes, películas o música
224

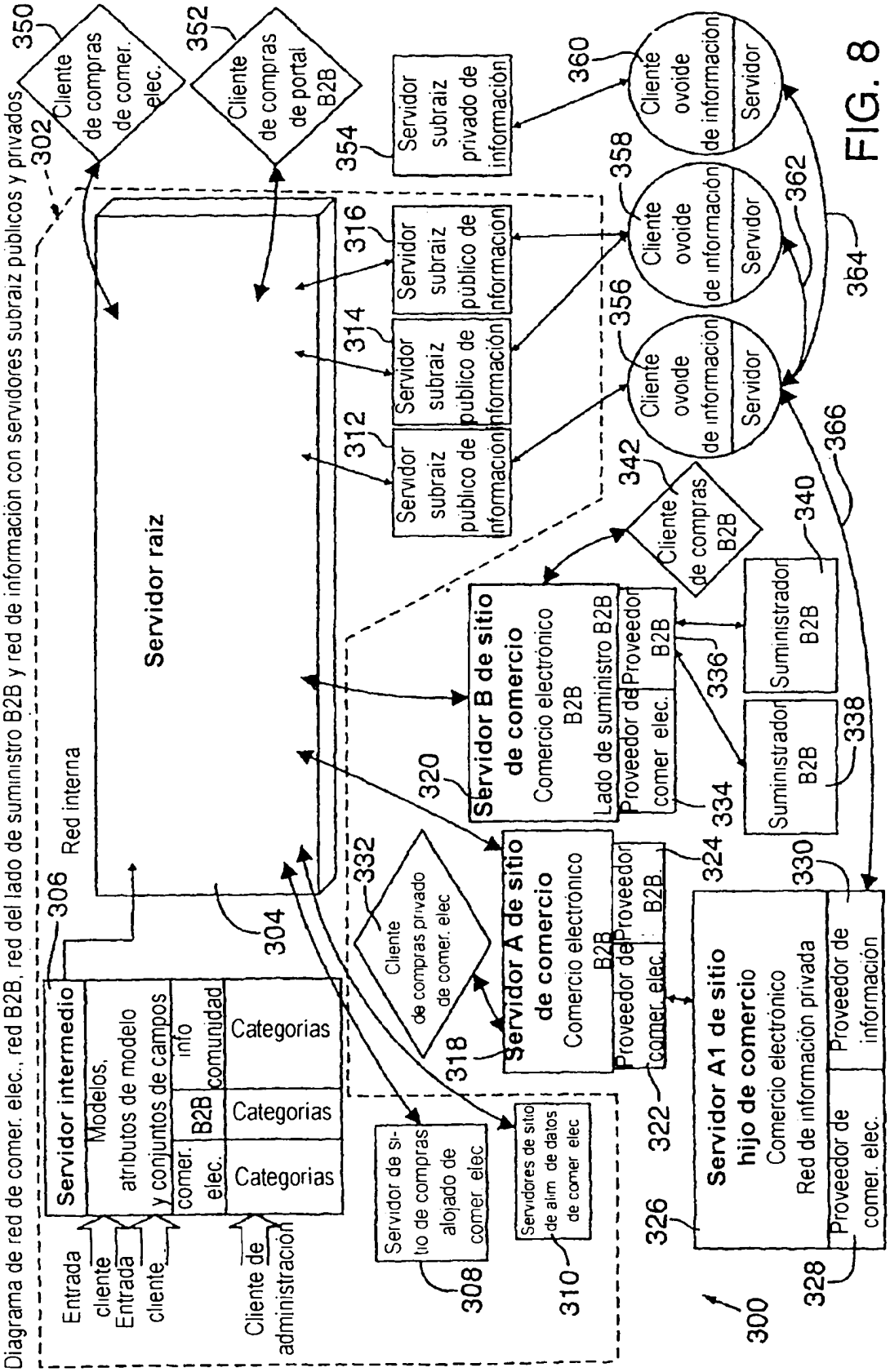
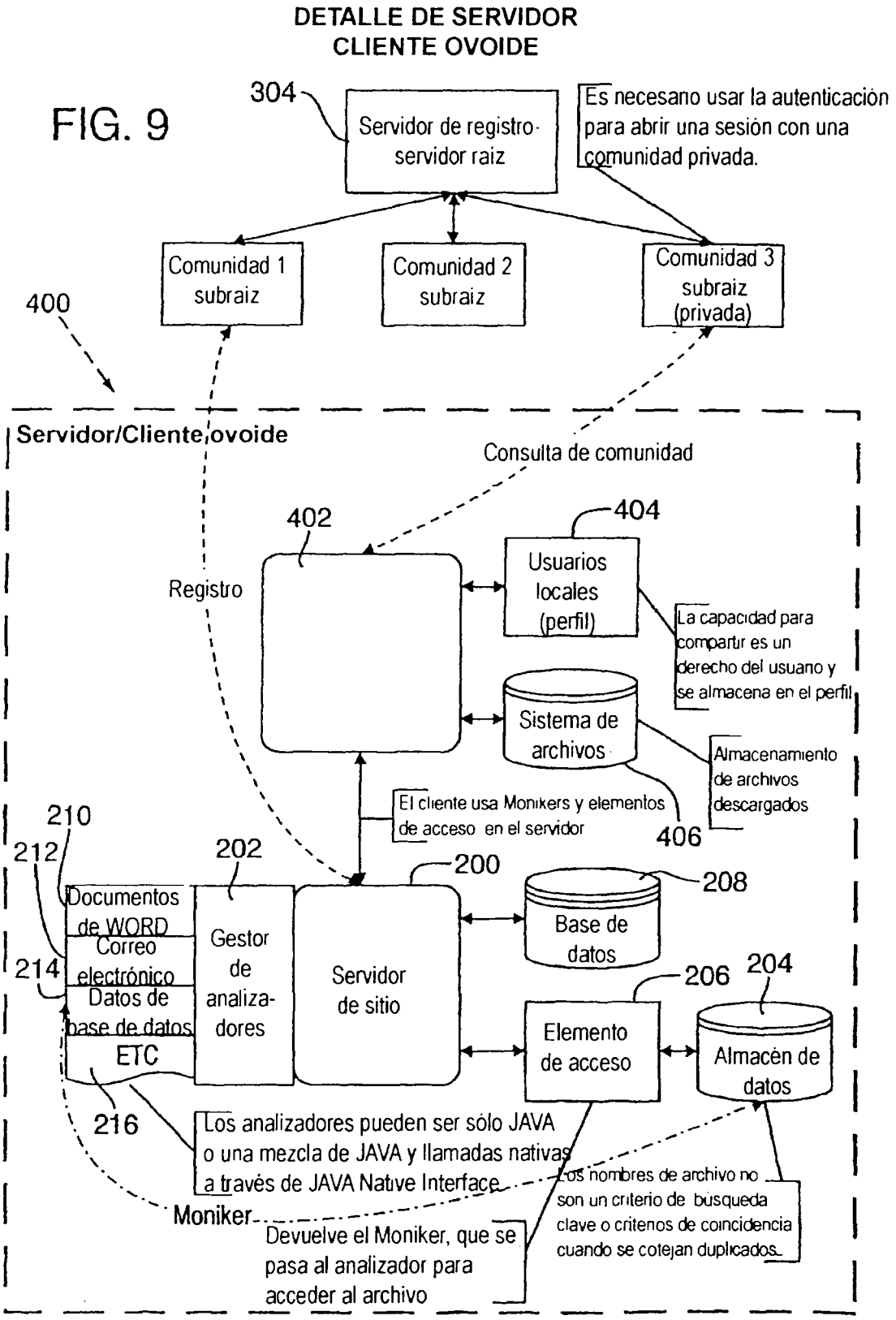


Diagrama de red de comer. elec., red B2B, red del lado de suministro B2B y red de información con servidores subraiz publicos y privados
Servidor raiz
Red interna
Servidor intermedio
Modelos, atributos de modelo y conjuntos de campos
comer. elec.
B2B
info comunidad
Categorias
Categorias
Categorias
Entrada cliente
Entrada cliente
Cliente de administración
Servidor de sitio de compras alojado de comer. elec
Servidores de sitio de alm. de datos de comer elec
Cliente de compras privado de comer. elec
Servidor A de sitio de comercio
Comercio electrónico B2B
Proveedor de comer. elec.
Proveedor B2B
Servidor B de sitio de comercio
Comercio electrónico B2B
Lado de suministro B2B
Proveedor de comer elec.
Proveedor B2B
Servidor A1 de sitio hijo de comercio
Comercio electrónico
Red de información privada
Proveedor de comer. elec.
Proveedor de información
Suministrador B2B
Suministrador B2B
Cliente de compras B2B
Cliente de compras de comer. elec.
Cliente de compras de portal B2B
Servidor subraiz privado de información
Servidor subraiz público de información
Servidor subraiz público de información
Servidor subraiz público de información
Cliente ovoide de información
Servidor
Cliente ovoide de información
Servidor
Cliente ovoide de información
Servidor
300
302
304
306
308
310
312
314
316
318
320
322
324
326
328
330
332
334
336
338
340
342
350
352
354
356
358
360
362
364
366

FIG. 8

FIG. 9


DETALLE DE SERVIDOR
CLIENTE OVOIDE
304
Servidor de registro-servidor raiz
Es necesario usar la autenticación para abrir una sesión con una comunidad privada.
Comunidad 1 subraiz
Comunidad 2 subraiz
Comunidad 3 subraiz (privada)
400
Servidor/Cliente ovoide
Consulta de comunidad
402
404
Registro
Usuarios locales (perfil)
La capacidad para compartir es un derecho del usuario y se almacena en el perfil
Sistema de archivos
Almacenamiento de archivos descargados
El cliente usa Monikers y elementos de acceso en el servidor
406
210
212
202
200
208
Documentos de WORD
Correo electrónico
214
Datos de base de datos
ETC
Gestor de analizadores
Servidor de sitio
Base de datos
206
204
Elemento de acceso
Almacén de datos
216
Los analizadores pueden ser sólo JAVA o una mezcla de JAVA y llamadas nativas a través de JAVA Native Interface
Moniker
Devuelve el Moniker, que se pasa al analizador para acceder al archivo
Los nombres de archivo no son un criterio de búsqueda clave o criterios de coincidencia cuando se cotejan duplicados

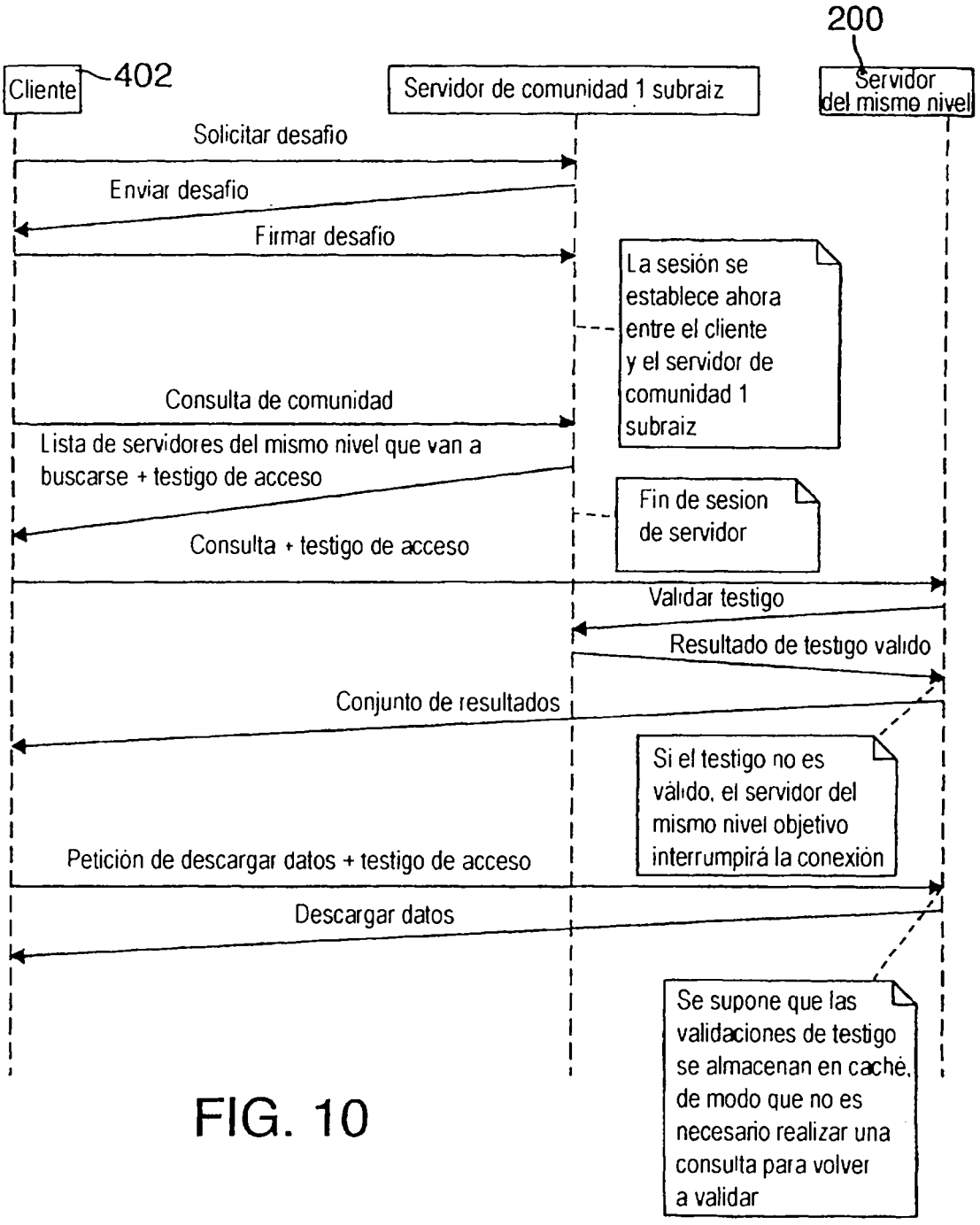

200
Cliente
402
Servidor de comunidad 1 subraiz
Servidor del mismo nivel
Solicitar desafio
Enviar desafio
Firmar desafio
La sesión se establece ahora entre el cliente y el servidor de comunidad 1 subraiz
Consulta de comunidad
Lista de servidores del mismo nivel que van a buscarse + testigo de acceso
Fin de sesion de servidor
Consulta + testigo de acceso
Validar testigo
Resultado de testigo valido
Conjunto de resultados
Si el testigo no es válido, el servidor del mismo nivel objetivo interrumpirá la conexión
Petición de descargar datos + testigo de acceso
Descargar datos
Se supone que las validaciones de testigo se almacenan en caché, de modo que no es necesario realizar una consulta para volver a validar

FIG. 10

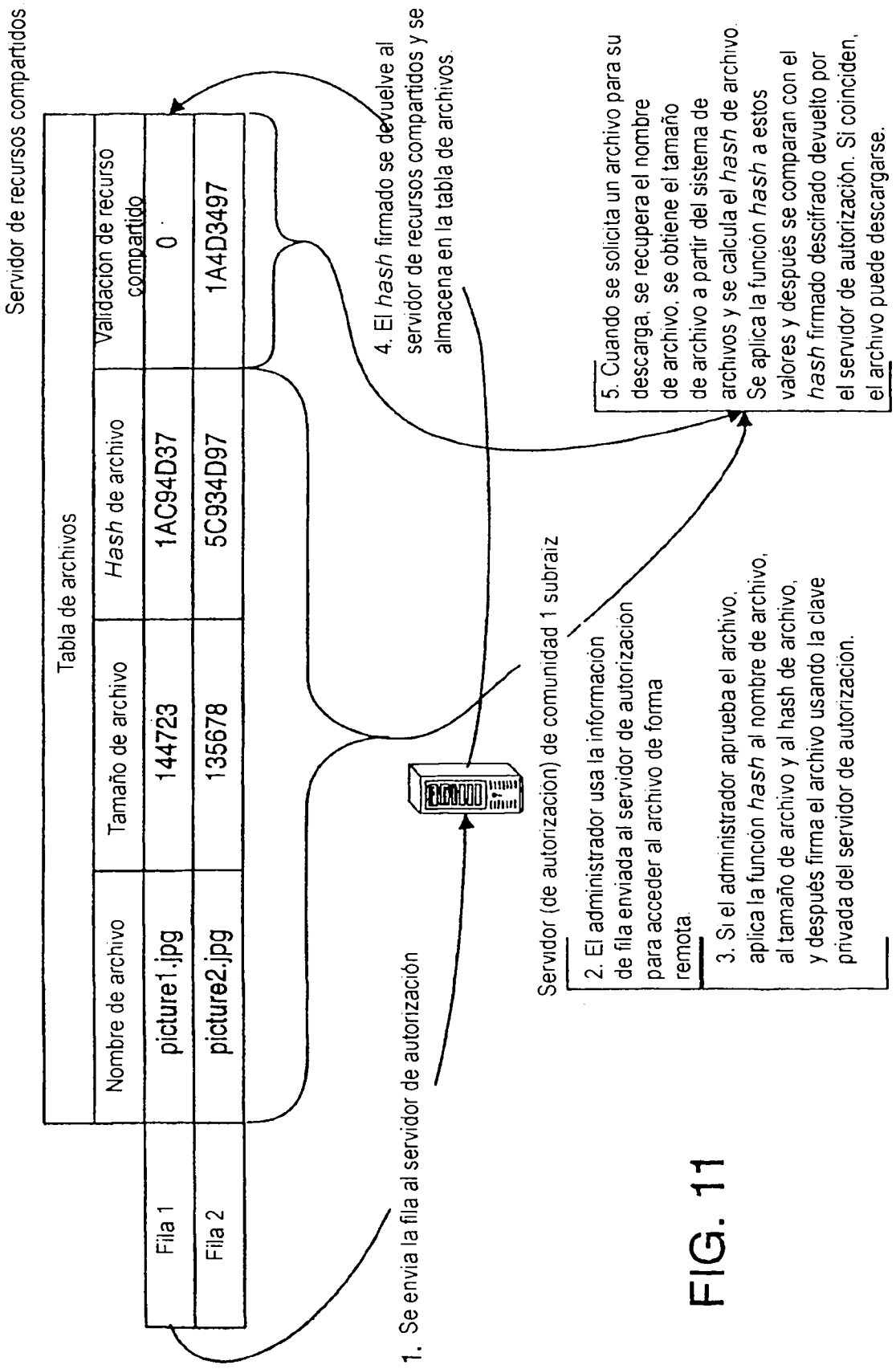

Servidor de recursos compartidos
Tabla de archivos
Nombre de archivo
Tamaño de archivo
Hash de archivo
Validación de recurso compartido
Fila 1
picture1.jpg
144723
1AC94D37
0
Fila 2
picture2.jpg
135678
5C934D97
1A4D3497
1. Se envía la fila al servidor de autorización
Servidor (de autorización) de comunidad 1 subraiz
2. El administrador usa la información de fila enviada al servidor de autorización para acceder al archivo de forma remota.
3. Si el administrador aprueba el archivo, aplica la función hash al nombre de archivo, al tamaño de archivo y al hash de archivo, y después firma el archivo usando la clave privada del servidor de autorización.
4. El hash firmado se devuelve al servidor de recursos compartidos y se almacena en la tabla de archivos.
5. Cuando se solicita un archivo para su descarga, se recupera el nombre de archivo, se obtiene el tamaño de archivo a partir del sistema de archivos y se calcula el hash de archivo. Se aplica la función hash a estos valores y después se comparan con el hash firmado descifrado devuelto por el servidor de autorización. Si coinciden, el archivo puede descargarse.

FIG. 11