(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 253 531**

(51) Int. Cl.[7]: **G06F 3/06**

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Número de solicitud europea: **02727013 .1**
(86) Fecha de presentación : **03.01.2002**
(87) Número de publicación de la solicitud: **1374028**
(87) Fecha de publicación de la solicitud: **02.01.2004**

(54) Título: **Gestión del almacenamiento sobre la base del contenido.**

(30) Prioridad: **03.01.2001 US 259158 P**

(45) Fecha de publicación de la mención BOPI:
01.06.2006

(45) Fecha de la publicación del folleto de la patente:
01.06.2006

(73) Titular/es: **Nice Systems Limited**
8 Hapnina Street, P.O. Box 690
Ra'anana 43107, IL

(72) Inventor/es: **Kasmirsky, Yehoshaphat y**
Kor, Ilan

(74) Agente: **Isern Jara, Jorge**

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

Gestión del almacenamiento sobre la base del contenido.

Ámbito de la invención

La presente invención se refiere a un sistema y a un procedimiento para la gestión del almacenamiento sobre la base del contenido y en particular a un sistema y a un procedimiento de tal tipo en el cual las decisiones relativas a la localización y el tiempo del almacenamiento de los datos se basa en el contenido de los datos.

Antecedentes de la invención

Las instalaciones de almacenamiento para la información digital son un recurso crítico. La demanda de espacio de almacenamiento tanto para datos convencionales, como por ejemplo documentos de texto y otros archivos legibles por el ser humano, como de flujos multimedia, tal como por ejemplo datos de audio y de vídeo, ha incrementado significativamente. Un incremento de este tipo resulta de una serie de factores diferentes, tales como los requisitos legales de almacenar y mantener ciertos tipos de información; un incremento de los diferentes tipos de datos que se almacenan; e incluso un incremento del tamaño de las unidades individuales de datos, tales como por ejemplo archivos de documentos de procesamiento de texto, archivos de datos de vídeo etcétera. Esta demanda creciente a su vez resulta en una mayor demanda de espacio de almacenamiento y en particular de espacio de almacenamiento que sea accesible "en línea".

A medida que la demanda de espacio de almacenamiento en línea aumenta, son posibles una serie de opciones para cubrir esa demanda. Por ejemplo, equipos adicionales, tales como dispositivos de medios magnéticos ("unidades de disco duro"), se pueden comprar para aumentar la cantidad disponible de espacio de almacenamiento electrónicamente accesible. Sin embargo, a medida que aumenta la cantidad de dispositivos de equipos de este tipo, el problema de la gestión para la gestión electrónica de tales dispositivos también aumenta. Adicionalmente, incrementar meramente el espacio de almacenamiento puede ser tanto inútil como innecesario, puesto que no todos los datos pueden requerir, o por lo menos no requerirán un acceso inmediato.

El problema se puede solucionar parcialmente a través de la utilización de una mezcla de tipos diferentes de instalaciones de almacenamiento. Por ejemplo, el almacenamiento en línea se refiere a áreas de almacenamiento permanentemente montadas de acceso directo, tales como unidades de disco magnético (o bien otros tipos de medios) y pilas de discos. El tiempo requerido para hacer el acceso a tales áreas de almacenamiento típicamente se mide en fracciones de segundo. Puesto que no todos los datos pueden necesitar ser almacenados en almacenamiento en línea, lo cual es rápido pero también caro, también está disponible el almacenamiento próximo a línea. El almacenamiento próximo a línea se basa en un área de almacenamiento automáticamente (máquina) accionada, tal como por ejemplo discos ópticos que residen en una "memoria con discos" o cintas en una biblioteca automática de cintas. Tales dispositivos de almacenamiento automáticamente accionados son capaces de almacenar y acceder automáticamente a una cantidad relativamente grande de datos con menos dispositivos de lectura física, o unidades, para leer los datos. Este tipo de almacenamiento es menos caro, pero también algo más lento en el acceso a los datos, de forma que los tiempos de acceso se miden desde segundos hasta minutos, o incluso más largos, dependiendo de la disponibilidad de las unidades físicas para la lectura de los medios almacenados. El almacenamiento en línea o próximo a línea también puede ofrecer un sistema con una pluralidad de unidades físicas, conectadas juntas, por ejemplo en una red de área local (LAN - local area network) o en una red de área extendida (WAN - wide area network).

El almacenamiento fuera de línea es el tipo de almacenamiento menos caro, pero también es el más lento de acceder y no permite el acceso electrónico automático. En cambio, se requiere un accionamiento manual por parte de un operario de los dispositivos de almacenamiento y de las unidades físicas. El número de unidades físicas se reduce enormemente comparado con el número de dispositivos de almacenamiento (o por lo menos la cantidad de espacio de almacenamiento disponible). Sin embargo, el tiempo de acceso a los datos desde tales dispositivos se mide desde minutos hasta horas, dependiendo de la disponibilidad del operario humano y de la localización de los dispositivos de almacenamiento, así como de la disponibilidad de las unidades físicas.

También se pueden utilizar otros tipos de dispositivos de almacenamiento y de funciones, además de, o el lugar de, los dispositivos y de funciones anteriormente descritos. En cualquier caso, la dificultad de un sistema mixto, o un sistema en el cual se utilizan diferentes tipos de áreas (topología) de almacenamiento, con diferentes tipos de dispositivos de almacenamiento y diferente accesibilidad (particularmente con respecto al tiempo de acceso), es la gestión de los datos. Ciertos tipos de datos pueden ser más importantes, o por lo menos más críticos desde el punto de vista del tiempo de acceso, de tal forma que el tiempo de acceso puede ser muy importante para algunos tipos de datos y mucho menos importante para otros tipos de datos. El coste es también un factor importante. Además, las decisiones se deben tomar teniendo en cuenta el número y el tipo de dispositivos de almacenamiento que se vayan a comprar, junto con cualquier dispositivo de soporte y soporte del sistema, tal como los operarios por ejemplo. Actualmente, estos sistemas están diseñados y construidos manualmente y las decisiones se realizan sobre la base de algún tipo de política. Sin embargo, el funcionamiento del sistema real e incluso el propio diseño puede no ser óptimo para una organización particular.

El documento EP-A-0950965 se refiere a un sistema y un procedimiento de gestión del almacenamiento de documentos que gestiona el almacenamiento de documentos sobre la base de la similitud del contenido de los documentos. Grupos de documentos se crean sobre la base de la similitud de los contenidos de los documentos. El almacenamiento del grupo o el documento seleccionado se gestiona entonces, por ejemplo, borrando, comprimiendo o copiando.

El documento de Hiroshi Ishikawa y otros: un almacén de documentos: enfoque de una base de datos multimedia. DEXA Workshop 1998 páginas 90-94 se refiere a un almacén de documentos como por ejemplo un marco de programas en el que los documentos multimedia son analizados y gestionados para compartir y reutilizar la información corporativa como un almacén de datos para datos estructurados. Describe un sistema de almacén de documentos prototipo, el cual soporta la gestión de documentos simples y compuestos, recuperación a partir de claves y a partir de contenidos, clasificación a partir de reglas y agrupamiento a partir del inicio de mensaje y reglas de gestión.

El documento EP-A-0341230 se refiere a un procedimiento para la gestión de datos en una jerarquía de almacenamiento de datos y una jerarquía de almacenamiento de datos para la implantación de un procedimiento de este tipo. La jerarquía de almacenamiento de datos incluye un catálogo de datos y una pluralidad de niveles de almacenamiento de datos. El catálogo de datos almacena la información indicando las reglas de la gestión de los datos para los datos almacenados en la jerarquía. Los datos almacenados en la jerarquía son señalados al ser referenciados mediante un procesador principal. Los datos señalados se gestionan de acuerdo con las reglas de gestión durante períodos de baja actividad jerárquica.

El documento US-A-5131087 se refiere a un aparato que incluye una serie de bases de datos las cuales son utilizadas por los programas del sistema experto para gestionar los dispositivos de almacenamiento de datos del sistema de ordenador. Un elemento provisto en este aparato es un conjunto de datos de configuración del dispositivo de almacenamiento de datos que proporciona una descripción de los diversos dispositivos de almacenamiento de datos y sus interconexiones en el sistema de ordenador. Un segundo elemento es una base de datos de conocimiento que incluye un conjunto de reglas funcionales que describen la función de gestión del dispositivo de almacenamiento de datos. Estas reglas indican las características funcionales de los diversos dispositivos de almacenamiento de datos.

El documento US-A-5506986 se refiere a un sistema de gestión de medios que utiliza un sistema a partir de reglas para gestionar la función de asignación de medios para un procesador de datos. Este aparato comprende un conjunto de programas de software que se comunican en tiempo real con diversos componentes de programas normales en el procesador de datos y también recuperan información que describe los datos almacenados en los propios dispositivos de almacenamiento de datos para compilar la información concerniente a la actividad del conjunto de datos.

Resumen de la invención

La técnica anterior no indica ni sugiere una solución al problema de gestionar eficazmente el almacenamiento de datos. La técnica anterior tampoco indica ni sugiere una solución al problema de la gestión del almacenamiento de datos tanto con relación a un buen rendimiento del coste como con relación a los tiempos acceso adecuados. La técnica anterior tampoco proporciona una solución para el almacenamiento de datos de acuerdo con el contenido de los datos, de forma que los datos importantes puedan ser almacenados en una ubicación o un tipo más accesible de almacenamiento de archivos. Además, la técnica anterior no indica ni sugiere un sistema y procedimiento para la gestión de datos, de forma que tales datos sean correctamente almacenados, puedan migrar y ser borrados, de acuerdo con el contenido de los mismos.

La presente invención soluciona estos problemas de la técnica anterior proporcionando un sistema y un procedimiento para la gestión de datos de acuerdo con el contenido de los datos. La presente invención capacita que los datos sean almacenados en una pluralidad de opciones de almacenamiento diferentes de acuerdo con por lo menos una característica de los datos, en la que la por lo menos una característica está relacionada con el contenido de los datos. Debe indicarse que por lo menos una opción de almacenamiento de acuerdo con la presente invención es opcionalmente el borrado y la destrucción de los datos, de tal forma que los datos pueden ser opcionalmente eliminados del medio de almacenamiento o bien opcionalmente pueden no ser almacenados inicialmente en el medio de almacenamiento. Por lo tanto, una “opción de almacenamiento” de acuerdo con la presente invención incluye cualquier tipo de medio, dispositivo, sistema de almacenamiento o combinación de los mismos, o borrado (eliminación) de los datos.

Opcionalmente y más preferiblemente, los datos se almacenan durante un intervalo de tiempo de acuerdo con la por lo menos una característica de los datos. Más preferiblemente, los datos se desplazan a un tipo diferente de opción de almacenamiento después de que haya ocurrido un evento, por ejemplo después de que haya pasado un intervalo de tiempo. Debe indicarse que el desplazamiento o migración de los datos puede incluir también el borrado o eliminación de los datos.

Breve descripción de los dibujos

La invención se describe aquí, a título de ejemplo únicamente, con referencia a los dibujos que se acompañan en los cuales:

la figura 1 es un diagrama de bloques esquemático de un sistema y un flujo de operaciones ejemplares de acuerdo con la presente invención;

la figura 2 es un diagrama de bloques esquemático de otro sistema y flujo de operaciones ejemplares de acuerdo con la presente invención; y

la figura 3 es un diagrama de bloques esquemáticos de un sistema ejemplar detallado de acuerdo con la presente invención.

Descripción de las realizaciones preferidas

La presente invención es un sistema y un procedimiento para la gestión de datos de acuerdo con el contenido de los datos. La presente invención capacita que los datos sean almacenados en una de una pluralidad de opciones de almacenamiento diferentes de acuerdo con por lo menos una característica de los datos, en la que la por lo menos una característica está relacionada con el contenido de los datos. Debe indicarse que por lo menos una opción de almacenamiento de acuerdo con la presente invención es opcionalmente el borrado y la destrucción de los datos, de tal forma que los datos pueden ser opcionalmente eliminados del medio de almacenamiento o bien opcionalmente pueden no ser almacenados inicialmente en el medio de almacenamiento. Opcionalmente y más preferiblemente, los datos se almacenan durante un intervalo de tiempo de acuerdo con la por lo menos una característica de los datos. Más preferiblemente, los datos se desplazan a un tipo diferente de opción de almacenamiento después de que haya ocurrido un evento, por ejemplo después de que haya pasado un intervalo de tiempo. Debe indicarse que el desplazamiento o migración de los datos puede incluir también el borrado o eliminación de los datos.

De acuerdo con las realizaciones preferidas de la presente invención, la por lo menos una característica de los datos, de acuerdo con la cual se selecciona la opción de almacenamiento, es examinada por un motor de reglas. Preferiblemente, el motor de reglas compara la por lo menos una característica de los datos con por lo menos una regla y selecciona entonces la opción (o las opciones) de almacenamiento de acuerdo con esa regla. El motor de reglas por lo tanto funciona más preferiblemente como un filtro, para determinar qué opción (u opciones) de almacenamiento es la más apropiada para los datos examinados. La decisión de almacenamiento es entonces implantada preferiblemente por un gestor de almacenamiento.

Las reglas de acuerdo con las cuales funciona el motor de reglas son opcionalmente introducidas manualmente por un operario, o alternativamente pueden ser opcionalmente generadas automáticamente de acuerdo con una regla de gestión previamente definida o de acuerdo con una regla de gestión automáticamente generada, o una combinación de las mismas.

Preferiblemente, la presente invención es operativa con un sistema que ofrece una pluralidad de opciones de almacenamiento diferentes. Más preferiblemente, estas diferentes opciones de almacenamiento incluyen por lo menos dos opciones almacenamiento diferentes provistas de tipos diferentes de accesibilidad. Ejemplos de opciones almacenamiento provistas de diferentes tipos de accesibilidad incluyen pero no están limitadas al almacenamiento en línea, el almacenamiento próximo a línea y el almacenamiento fuera de línea. El tipo de medio de almacenamiento que se utiliza para cualquier opción de almacenamiento particular no está limitado de acuerdo con la presente invención, ya que la presente invención es operativa con cualquier tipo adecuado de medio de almacenamiento, incluyendo pero no está limitado al almacenamiento a partir de cintas (DAT), también el almacenamiento a partir de cintas (AIT), medios almacenamiento magnético, discos ópticos, CD-ROM o dispositivos de almacenamiento masivo de cualquier tipo, o cualquier tipo de sistema de almacenamiento o cualquier combinación de los mismos.

La por lo menos una característica de los datos, la cual está relacionada con el contenido de los datos, se puede obtener por una serie de modos diferentes. Por ejemplo, los datos pueden tener asociados opcionalmente y preferiblemente metadatos, los cuales están relacionados con el contenido de los datos. Los metadatos preferiblemente se añaden a través de una anotación a los propios datos. Tales anotaciones se llevan a cabo automáticamente. Más preferiblemente, la anotación automática se lleva a cabo después de que los datos hayan sido analizados automáticamente. Los metadatos asociados son entonces utilizados preferiblemente para determinar qué opción de almacenamiento se debe utilizar para los datos y más preferiblemente también el intervalo de tiempo durante el cual los datos se debe colocar en esa opción de almacenamiento.

Como se ha descrito anteriormente, los datos son preferiblemente filtrados por un motor de reglas, de acuerdo con por lo menos una característica de los datos. Para la implantación de la presente invención con metadatos, el proceso de filtrado se lleva a cabo más preferiblemente de acuerdo con los metadatos asociados.

El análisis automático de los datos se lleva a cabo más preferiblemente de acuerdo con el tipo de datos que están siendo analizados. Ejemplos de diferentes tipos de procesos de análisis automático los cuales se puede llevar a cabo opcionalmente incluyen pero no están limitados a la conversión de dictado a texto para los datos de comunicación de voz, un analizador de vídeo para los datos de vídeo, un lector óptico de caracteres para el material impreso que haya sido rastreado electrónicamente, análisis de imágenes para datos de imágenes, análisis de texto para datos de texto y analizadores para datos de interfaz del usuario. Estos tipos diferentes de procesos de análisis de datos se llevan a cabo preferiblemente de acuerdo con la fuente de los datos, la cual puede ser opcionalmente cualquier fuente de datos adecuada. Ejemplos de diferentes tipos de fuentes de datos incluyen pero no están limitados a datos de vídeo, datos de audio (incluyendo también datos de comunicación por voz tales como datos de voz sobre protocolo de internet (IP - internet protocol) (VoIP), datos de flujo de audio y cualquier otro tipo de datos relacionados con audio), datos

codificados, mensajes de correo electrónico y anexos, conversaciones y otros tipos de mensajes de un sistema de mensajería, documentos transmitidos por facsímil y datos de la interfaz del usuario.

Además, la presente invención es útil para recoger datos sobre sustancialmente cualquier tipo de función de interfaz del usuario. Ejemplos de funciones de interfaz del usuario incluyen pero no están limitadas a cualquier tipo de actividades de ventana de la interfaz gráfica de usuario (GUI - graphical user interface), actividad con artilugios de la interfaz gráfica de usuario, botones, correderas o cualquier función provista a través de la ventana de la interfaz gráfica de usuario; la visualización de cualquier imagen o texto, incluye pero no está limitada a páginas de la red y cualquier componente de las mismas; información provista a través de una interfaz audible como por ejemplo voz sintetizada; información provista a través de la visualización de datos de vídeo; y cualquier otro tipo de información que sea proporcionada a través, o bien se pueda detectar de otro modo, del sistema operativo del dispositivo de ordenador del usuario.

De acuerdo con las realizaciones preferidas de la presente invención, los datos son preferiblemente "migrados" o desplazados desde una primera opción de almacenamiento a una segunda opción de almacenamiento diferente después de que haya expirado un intervalo de tiempo. El intervalo de tiempo preferiblemente se determina de acuerdo con los metadatos.

Los principios y el funcionamiento del procedimiento de acuerdo para presente invención se entenderán mejor con referencia a los dibujos y a la descripción que se acompañan.

Con referencia ahora a los dibujos, la figura 1 muestra un primer sistema ejemplar 10 para la gestión de los datos de acuerdo con el contenido de los datos, con respecto al flujo de funciones a través del sistema de 10. Como se representa, el sistema 10 está equipado con por lo menos una fuente de datos de entrada 12. Ejemplos de diferentes tipos de fuentes de datos de entrada 12 incluyen pero están limitados a datos de vídeo, datos de audio (incluyendo también datos de comunicación por voz tales como datos de voz sobre protocolo de internet (VoIP), datos de flujo de audio y cualquier otro tipo de datos relacionados con audio), datos codificados, mensajes de correo electrónico y anexos, conversaciones y otros tipos de mensajes de un sistema de mensajería, documentos transmitidos por facsímil y datos de la interfaz del usuario. Con respecto a los datos de la interfaz del usuario, en los cuales la acción de un usuario al accionar el dispositivo ordenador y el dispositivo periférico y el dispositivo de introducción del mismo causa que los datos sean generados, opcionalmente y preferiblemente los datos están en forma de un evento. Cada acción del usuario preferiblemente causa que se genere un evento. El evento puede entonces opcionalmente formar los datos que se van a capturar.

Los datos capturados se pasan a un analizador de formato 14 para transformar los datos capturados en un formato común para el análisis. El analizador de formato 14 preferiblemente está equipado con una pluralidad de módulos de formato 16, cada uno de los cuales es adecuado para datos de diferentes tipos de formato. Por ejemplo, si los datos de entrada son datos de comunicación de voz, entonces preferiblemente un módulo de formato 16 convierte la comunicación por voz en datos textuales, para una conversión de voz a texto. Diferentes módulos de formato 16 preferiblemente manejan otros tipos de datos de entrada, como se explica con mayor detalle con referencia a la figura 3 más adelante.

De acuerdo con una realización preferida de la presente invención, el formato de los datos común es opcionalmente y preferiblemente datos textuales. Para esta realización preferida, los datos textuales no son opcionalmente procesados adicionalmente por un módulo de formato 16, o alternativamente sólo son procesados mínimamente. Otros tipos de datos en formatos de datos diferentes son entonces convertidos preferiblemente en datos textuales mediante un módulo de formato 16, como se describe con respecto a la figura 3 más adelante.

A continuación un módulo de análisis 18 preferiblemente analiza los datos, una vez los datos están en el formato común. Debe indicarse que el módulo de análisis 18 opcionalmente es capaz de manejar una pluralidad de formatos de datos diferentes, a través de un único módulo o alternativamente a partir de un conjunto de tales módulos (no representados). Preferiblemente, el módulo de análisis 18 funciona en una pluralidad de diferentes tipos y fuentes de datos simultáneamente, por ejemplo como una aplicación de múltiples series. Para el ejemplo ilustrativo no limitativo, como se ha descrito antes, los datos textuales se analizan, como es muy conocido en la técnica. Un ejemplo no limitativo de un programa analizador de texto es el producto IntelligentMiner™ (IBM Corp, véase http://www-4.ibm.com/software/data/iminer/fortext/tatools.html de 31 de diciembre de 2001 para detalles adicionales). Este producto es útil para analizar texto para una serie de funciones tales como información de localización relativa a un tema, categorización de información y clasificación de información. Las herramientas analizadoras de texto generalmente son conocidas en la técnica para extraer el contenido y la información relativa al tema objeto del texto, por ejemplo de acuerdo con una o más palabras clave, conceptos o cualquier otro tipo de esquemas de organización y análisis.

De acuerdo con otra realización preferida de la presente invención, el formato uniforme ofrece una estructura de datos uniforme, con una pluralidad de tipos y categorías diferentes de información, por ejemplo datos a partir de eventos de pantalla y datos de voz combinados en un único archivo. Esta estructura de datos uniforme preferiblemente es capaz de contener tipos o categorías diferentes de características las cuales son de interés para ser asociadas con los datos, a fin de determinar el contenido de los datos capturados. Ejemplos no limitativos de una estructura de datos uniforme que puede ser opcionalmente implantada de acuerdo con la presente invención incluyen estructuras

que utilizan el lenguaje de marcas extensible (XML - extensible markup language) o formato de flujo (multimedia) avanzado (ASF - advanced streaming format).

El módulo de análisis 18 preferiblemente extrae o crea, o determina de otro modo, por lo menos una característica de los datos capturados, preferiblemente obtenidos en el formato uniforme. Opcionalmente y más preferiblemente, el módulo de análisis 18 obtiene la por lo menos una característica de los datos capturados en la forma de metadatos. Estos metadatos son entonces opcionalmente almacenados en una base de datos de metadatos (no representada, véase la figura 3).

De acuerdo con una realización opcional pero preferida de la presente invención, el módulo de análisis 18 también proporciona retroalimentación para mejorar el comportamiento del analizador de formato 14 y el módulo de formato 16, a fin de mejorar el funcionamiento de estos componentes.

Los datos capturados, con la por lo menos una característica relacionada con el contenido de los datos, los cuales tienen la forma opcionalmente y más preferiblemente de metadatos, son pasados entonces al motor de reglas 20. Alternativamente, el motor de reglas 20 sólo recibe la por lo menos una característica, más preferiblemente en forma de metadatos. Los datos capturados son pasados entonces directamente a un gestor de almacenamiento 22. El motor de reglas 20 más preferiblemente compara los metadatos con la por lo menos una regla, la cual es más preferiblemente una regla de gestión especificada por un usuario manual, o alternativamente puede ser opcionalmente generada automáticamente de acuerdo con una regla de gestión previamente definida o de acuerdo con una regla de gestión automáticamente generada, o una combinación de las mismas. Opcionalmente, una o más reglas puede ser alimentadas al motor de reglas 20 a través de una interfaz 24, tal como por ejemplo una interfaz gráfica de usuario por ejemplo. Como ejemplo no limitativo, la interfaz 24 puede ser opcionalmente una simple interfaz a partir del navegador de la red. El motor de reglas 20 determina entonces preferiblemente el tipo de opción (u opciones) de almacenamiento de acuerdo con una o más reglas, como se selecciona a través de la comparación de los metadatos (o característica) de los datos capturados con la regla (o las reglas). Adicionalmente o alternativamente, la salida del motor de reglas 20 es opcionalmente y preferiblemente retroalimentada al analizador de formato 14, al módulo de formato 16 y al módulo de análisis 18.

El motor de reglas 20 opcionalmente y más preferiblemente determina tanto el tipo de opción de almacenamiento (u opciones), las cuales deben ser seleccionadas para los datos capturados particulares, como también el plazo del almacenamiento. Más preferiblemente, los datos capturados son almacenados inicialmente con una primera opción de almacenamiento y entonces son migrados (desplazados) a por lo menos una opción de almacenamiento adicional después de que haya pasado un período de tiempo. Este período de tiempo está también preferiblemente determinado de acuerdo con por lo menos una regla mediante el motor de reglas 20.

El motor de reglas 20 puede llevar a cabo opcionalmente una acción de acuerdo con una regla o evento, en el que el evento puede opcionalmente y preferiblemente disparar una aplicación automática de la regla. Ejemplos de acciones basadas en reglas se proporcionan más adelante. Un ejemplo de las acciones las cuales se basan en reglas incluyen pero no están limitadas a manipulaciones de los datos almacenados. Por ejemplo, la compresión de los datos que puede opcionalmente ser alterada después de que haya ocurrido un evento, tal como por ejemplo que haya pasado un período de tiempo. Un ejemplo no limitativo de una razón para alterar una compresión de este tipo es capacitar una reproducción más rápida de los datos.

Los datos almacenados anteriormente pueden ser actualizados opcionalmente y preferiblemente con datos de gestión. Ejemplos no limitativos de tales datos incluyen la adición de un número de la seguridad social o de identificación, la identificación del cliente, información sobre el estatus de cliente preferido etcétera.

Los datos de transacción actual también pueden opcionalmente y preferiblemente estar conectados a un archivo de datos anteriormente almacenados. Los datos anteriormente almacenados también pueden ser opcionalmente actualizados. Un ejemplo no limitativo de una conexión de este tipo se puede llevar a cabo opcionalmente mediante la conexión de transacciones llevadas a cabo por un cliente de cierto estatus elevado o preferido a una transacción pasada por ese cliente.

Otro ejemplo de evento o acción puede ser opcionalmente llevado a cabo con múltiples réplicas de datos, por ejemplo distribuyendo datos idénticos a diversos destinos y localizaciones u opciones de almacenamiento previamente definidas. Tales réplicas pueden ser opcionalmente llevadas a cabo con fines de redundancia, por ejemplo para la seguridad de los datos almacenados, duplicando múltiples opciones o localizaciones de almacenamiento y por razones de disponibilidad general.

El gestor de almacenamiento 22 preferiblemente recibe entonces la salida del motor de reglas 20 y los datos capturados. La salida del motor de reglas 20 preferiblemente incluye por lo menos una opción almacenamiento para los datos capturados. Como se ha indicado anteriormente, esta opción de almacenamiento puede ser un tipo de medio de almacenamiento o el borrado o eliminación y destrucción de los datos capturados. Más preferiblemente, la opción almacenamiento incluye un dispositivo de almacenamiento particular 26 dentro del cual se colocarán los datos capturados. Alternativamente, el gestor de almacenamiento 22 puede determinar la identidad de un dispositivo de almacenamiento particular 26 para almacenar los datos.

La figura 2 muestra una disposición diferente del flujo del sistema de la figura 1, como otro ejemplo de un sistema de acuerdo con la presente invención. En esta disposición, el sistema 10 está equipado otra vez con un motor de reglas 40. Sin embargo, en esta implantación, el motor de reglas 40 es el proceso o componente de inicio para subsiguientes acciones que son llevadas a cabo por el sistema 10. Como se representa, el motor de reglas 40 toma la fuente de entrada de acuerdo con los metadatos obtenidos a partir del módulo de análisis 38. El motor de reglas 40 envía entonces preferiblemente los datos capturados, o alternativamente sólo los datos capturados seleccionados, al gestor de almacenamiento 42. El gestor de almacenamiento 42 envía los datos capturados a la opción de almacenamiento correcta, representada preferiblemente como un dispositivo de almacenamiento seleccionado 44, de acuerdo con una demanda de acción por parte del motor de reglas 40. Opcionalmente y más preferiblemente, y más preferiblemente que necesario, el motor de reglas 40 retroalimenta los datos capturados y la información sobre los datos capturados, dentro de las fuentes de entrada. Los metadatos mejorados puede ser opcionalmente obtenidos a partir del módulo de análisis 38.

De acuerdo con una implantación opcional de la presente invención, el usuario define una tarea en el motor de reglas 40 para conseguir ciertos tipos de información, tal como por ejemplo información sobre llamadas de teléfono específicas. El motor de reglas 40 utiliza entonces preferiblemente el módulo de análisis 38 para seleccionar los datos específicos. El módulo de análisis 38 puede estar opcionalmente implantado como un servidor de gestión de llamadas, por ejemplo. Los datos específicos seleccionados puede ser opcionalmente cualquiera de uno o más datos de voz, datos capturados a partir de acciones de la interfaz del usuario, datos de vídeo, una transacción de correo electrónico, datos de facsímil, datos de voz sobre protocolo de internet (VoIP), datos de navegación por la red (obtenidos a partir de dos o más usuarios que visualizan la misma página de la red a través de diferentes procesos de navegador de la red), o cualquier dato codificado o combinación de cualquier tipo de fuentes de entrada. Los datos se obtienen a partir de las fuentes entrada 36, las cuales pueden estar opcionalmente implantadas como un gestor de registro de actividades de fuentes de entrada. Los datos capturados son entonces transferidos al gestor de almacenamiento 42.

Estos datos pueden ser opcionalmente recuperados como se requiera de los dispositivos almacenamiento 44, más preferiblemente directamente utilizando el motor de reglas 40 y el gestor de almacenamiento 42. Tales datos recuperados pueden ser entonces opcionalmente y más preferiblemente alimentados en fuentes de entrada 36 a fin de que los datos recuperados sean ampliamente accesibles (disponibles) y para una manipulación adicional.

La figura 3 muestra una implantación ejemplar detallada de un sistema de acuerdo con la presente invención, el cual está relacionado con el flujo de operaciones representado en la figura 1. Como se representa, un sistema 50 preferiblemente está equipado con una pluralidad de fuentes de datos de entrada 52. Para el propósito de la explicación únicamente y sin ninguna intención de que sea limitativo, las fuentes de datos de entrada 52 se representan opcionalmente y preferiblemente incluyendo una fuente de vídeo 54 para datos de vídeo, una fuente de audio 56 para datos de audio, una fuente de mensajes 58 para mensajes de correo electrónico (incluyendo anexos), datos de mensajes instantáneos y de conversación y una fuente de facsímil 60 para datos que son transmitidos por facsímil.

Los datos de entrada de las fuentes de datos de entrada 52 son preferiblemente alimentados entonces a por lo menos un analizador de formato 62 para convertir los datos capturados en un formato común para el análisis. El analizador de formato 62 preferiblemente está equipado con una pluralidad de módulos de formato 74, cada uno de los cuales es adecuado para datos de un tipo diferente de formato. Con propósitos ilustrativos únicamente y sin intención alguna de que sea limitativo, los módulos de formato 74 se representan opcionalmente y preferiblemente incluyendo un analizador de vídeo 76, un analizador de texto 78, un analizador de audio 79 y un módulo lector óptico de caracteres (OCR - optical caracter recognition) 80.

Un ejemplo de analizador de texto 78 ha sido descrito anteriormente con respecto a la figura 1. El módulo lector óptico de caracteres 80 opcionalmente puede estar implantado como es muy conocido en la técnica, por ejemplo a través de la utilización de programas de lector óptico de caracteres provistos de un algoritmo el cual puede ser fácilmente seleccionado por una persona normalmente experta en la técnica.

El análisis del vídeo opcionalmente se puede llevar a cabo mediante el analizador de vídeo 76 como sigue. Los datos de vídeo son obtenidos, por ejemplo a partir de una cámara como un ejemplo no limitativo de fuente de vídeo 54. Un capturador de cuadros es utilizado entonces preferiblemente para obtener por lo menos un cuadro de los datos de vídeo. El cuadro preferiblemente se analiza. Más preferiblemente, sólo una parte del cuadro se almacena como los datos capturados. Por ejemplo, si se utiliza una cámara de vídeo para supervisar la entrada a un área de seguridad, entonces opcionalmente sólo son de interés aquellos cuadros, o alternativamente aquellas partes de cada cuadro, las cuales caracterizan a un ser humano cerca de la entrada real. Adicionalmente o alternativamente, los cambios en el fondo de cada cuadro pueden ser opcionalmente detectados y seguidos, si son de interés.

Un ejemplo de un tipo de análisis que puede ser llevado a cabo con los datos de vídeo es un algoritmo de detección del movimiento, el cual es muy conocido en la técnica. Otro ejemplo son los algoritmos para el reconocimiento de la cara, los cuales también son muy conocidos en la técnica. Ejemplos no limitativos de algoritmos de análisis de vídeo se describen en: http://www.cs.rochester.edu/u/nelson/research/motion/motion.html del 31 de diciembre de 2001, y en http://www-white.media.mit.edu/vismod/demos/facerec del 31 de diciembre de 2001 para algoritmos de reconocimiento de la cara. Más preferiblemente, tales análisis son llevados a cabo con firmware, tal como por ejemplo un procesador de señales digitales (DSP - digital signal processor). Los resultados entonces pueden ser opcionalmente almacenados como los datos capturados.

La salida del analizador de formato 62 preferiblemente ofrece por lo menos una característica de los datos capturados los cuales más preferiblemente son los metadatos, como se ha descrito anteriormente. Los metadatos son almacenados más preferiblemente en una base de datos de metadatos 82. Un motor de reglas 84 analiza entonces preferiblemente los metadatos a partir de la base de datos de metadatos 82 (o alternativamente obtenidos directamente del analizador de formato 62), a fin de aplicar una o más reglas a los datos capturados. El motor de reglas 84 opcionalmente puede estar implantado con el producto BlazeSoftware Advisor de Blaze Software (véase http://www.blazesoft.com/products del 31 de diciembre de 2001 para más detalles). El motor de reglas 84 también puede estar opcionalmente implantado como un motor de reglas/tareas a partir de Nice Systems Ltd. (Ra'anana, Israel), por ejemplo sobre la base de los datos de gestión. Sin tener en cuenta la implantación específica, el motor de reglas 84 preferiblemente funciona de acuerdo con por lo menos una regla de gestión.

El motor de reglas 84 preferiblemente compara la por lo menos una característica de los datos con por lo menos una regla y entonces selecciona la opción (o las opciones) de almacenamiento de acuerdo con esa regla. El motor de reglas 84 funciona por lo tanto más preferiblemente como un filtro, para la determinación de qué opción (u opciones) de almacenamiento es la más apropiada para los datos examinados. La decisión de almacenamiento es entonces preferiblemente implementada mediante un gestor de almacenamiento 86.

Preferiblemente, el gestor de almacenamiento 86 es capaz de seleccionar a partir de una pluralidad de diferentes opciones de almacenamiento. Más preferiblemente, estas diferentes opciones de almacenamiento incluyen por lo menos dos opciones de almacenamiento diferentes provistas de diferentes tipos de accesibilidad. Ejemplos de opciones de almacenamiento provistas de diferentes tipos de accesibilidad incluyen pero no están limitadas al almacenamiento en línea, el almacenamiento próximo a línea y el almacenamiento fuera del línea. El tipo de medio de almacenamiento que se utiliza para cualquier opción particular de almacenamiento no está limitado de acuerdo con la presente invención. Para el propósito de la ilustración sólo y sin intención alguna de que sea limitativo, el gestor de almacenamiento 86 se representa capaz de seleccionar a partir de una pluralidad de dispositivos de almacenamiento, representado como un dispositivo de almacenamiento en línea 88 y un dispositivo de almacenamiento fuera del línea 90. Por supuesto, se pueden utilizar otros tipos de dispositivos y sistemas de almacenamiento en lugar, o además, de estos ejemplos de dispositivos de almacenamiento.

El motor de reglas 84 opcionalmente y preferiblemente es capaz de retroalimentar información al analizador de formato 62 para mejorar el comportamiento del analizador de formato 62.

Un ejemplo no limitativo de funcionamiento del sistema 50 se puede llevar a cabo como sigue. El sistema 50 puede estar opcionalmente implantado en un centro de servicios el cual procesa los servicios requeridos remotamente por los clientes, de forma que el cliente no está físicamente presente en el centro de servicios. El cliente por lo tanto entra en contacto con el personal del centro de servicios, por ejemplo a través de una comunicación por voz (tal como una llamada por teléfono por ejemplo), mensajes de correo electrónico, facsímiles etcétera. Una pluralidad de reglas de gestión han sido definidas e implantadas por motor de reglas 84, las cuales pueden incluir opcionalmente las siguientes reglas: se guarda una grabación para cada contacto del cliente que se refiera a una transacción financiera durante por lo menos tres meses incluso aunque no haya ocurrido transacción y se guarda para cada contacto en los cuales haya ocurrido una transacción financiera durante por lo menos siete años. Además, la grabación para cada contacto resultante en una transacción financiera real es almacenada primero en el almacenamiento en línea durante un mes y entonces en un almacenamiento fuera de línea durante el resto del plazo de siete años.

Una vez el cliente ha entrado en contacto con el operador del centro de servicios, por ejemplo a través del teléfono, se proporcionan datos a través de una fuente de audio 56 como un ejemplo de fuente de entrada 52. Estos datos de audio capturados son analizados, por ejemplo a fin de determinar si la transacción financiera ha ocurrido durante el contacto. Si ha ocurrido una transacción de este tipo, entonces los metadatos asociados con los datos de audio capturados indican una ocurrencia de este tipo. El analizador de formato 62 y particularmente el analizador de audio 72 preferiblemente analiza los datos de audio capturados para obtener tales metadatos.

Los propios datos a partir de la llamada son preferiblemente manipulados de acuerdo con una o más reglas de gestión, las cuales tanto pueden estar opcionalmente definidas manualmente como generadas automáticamente, a través del funcionamiento del motor de reglas 84. Preferiblemente, el motor de reglas 84 genera entonces una acción para ser llevada a cabo por el gestor de almacenamiento 86. Por ejemplo, el gestor de almacenamiento 86 puede opcionalmente almacenar los datos a partir de la llamada, migrar los datos a un nuevo tipo de almacenamiento, o borrar los datos, o bien cualquier otra acción o combinación de acciones las cuales pueden ser llevadas a cabo de acuerdo con uno o más eventos. Por lo tanto, el motor de reglas 84 es capaz de generar una o más instrucciones para la ejecución por el gestor de almacenamiento 86.

Mientras la invención ha sido descrita con respecto a un número limitado de realizaciones, se apreciará que se pueden realizar muchas variaciones, modificaciones y otras aplicaciones de la invención, dentro del ámbito de las reivindicaciones anexas.

REIVINDICACIONES

1. Procedimiento para la gestión del almacenamiento de datos de acuerdo con el contenido de los datos obtenidos a partir de una o más fuentes de entrada, comprendiendo:

- convertir los datos a un formato común y subsiguientemente analizar automáticamente los datos en dicho formato común a fin de determinar automáticamente la por lo menos una característica de los datos de acuerdo con el contenido y de acuerdo con dicha por lo menos una característica TANTO seleccionando uno de una pluralidad de medios de almacenamiento, dispositivo de almacenamiento, sistema de almacenamiento o una combinación de los mismos y colocando los datos dentro o eliminando los datos de dicho medio de almacenamiento seleccionado, dispositivo de almacenamiento, sistema de almacenamiento o combinación de los mismos; COMO seleccionando no almacenar los datos.

2. El procedimiento de la reivindicación 1 en el que dicha pluralidad de medios de almacenamiento, dispositivo de almacenamiento, sistema de almacenamiento o una combinación de los mismos tiene por lo menos dos tipos diferentes de dispositivos.

3. El procedimiento de la reivindicación 2 en el que dichos tipos diferentes de dispositivos tienen diferentes tiempos de acceso.

4. El procedimiento de cualquiera de las reivindicaciones 1-3 en el que por lo menos un medio de almacenamiento, dispositivo de almacenamiento, sistema de almacenamiento o una combinación de los mismos comprende un dispositivo de almacenamiento en línea.

5. El procedimiento de cualquiera de las reivindicaciones 1-4 en el que por lo menos un medio de almacenamiento, dispositivo de almacenamiento, sistema de almacenamiento o una combinación de los mismos comprende un dispositivo de almacenamiento fuera de línea.

6. El procedimiento de cualquiera de las reivindicaciones 1-5 en el que por lo menos un medio de almacenamiento, dispositivo de almacenamiento, sistema de almacenamiento o una combinación de los mismos comprende un dispositivo de almacenamiento próximo a línea.

7. El procedimiento de cualquiera de las reivindicaciones 1-6 en el que dicha por lo menos una característica de los datos incluye metadatos asociados con los datos.

8. El procedimiento de la reivindicación 7 en el que dichos metadatos se obtienen analizando los datos.

9. El procedimiento de la reivindicación 8 en el que los datos incluyen una pluralidad de diferentes tipos de datos y dicha pluralidad de diferentes tipos de datos se analizan concurrentemente.

10. El procedimiento de cualquiera de las reivindicaciones 1-9 en el que dicha por lo menos una característica de los datos se compara con por lo menos una regla para la determinación de dicho medio de almacenamiento, dispositivo de almacenamiento, sistema de almacenamiento o una combinación de los mismos que se va a seleccionar.

11. El procedimiento de la reivindicación 10 en el que dicha por lo menos una regla incluye un intervalo de tiempo para mantener los datos en dicho medio de almacenamiento, dispositivo de almacenamiento, sistema de almacenamiento o una combinación de los mismos seleccionado.

12. El procedimiento de la reivindicación 11 en el que los datos son migrados desde un primer medio de almacenamiento, dispositivo de almacenamiento, sistema de almacenamiento o combinación de los mismos seleccionado a un segundo medio de almacenamiento, dispositivo de almacenamiento, sistema de almacenamiento o combinación de los mismos seleccionado después de que haya pasado dicho intervalo de tiempo.

13. El procedimiento de cualquiera de las reivindicaciones 10-12 en el que por lo menos una regla se introduce manualmente.

14. El procedimiento de cualquiera de las reivindicaciones 10-12 en el que dicha por lo menos una regla se genera automáticamente.

15. El procedimiento de la reivindicación 14 en el que dicha por lo menos una regla se genera automáticamente de acuerdo con los datos de gestión.

16. El procedimiento de cualquiera de las reivindicaciones 10-15 en el que dicha por lo menos una regla incluye una acción para ser llevada a cabo en los datos de acuerdo con un evento, en el que dicho evento está relacionado con dicha por lo menos una característica de los datos.

17. El procedimiento de cualquiera de las reivindicaciones 1-16 en el que dichos datos a partir de dicha fuente de entrada incluye por lo menos uno de los datos de vídeo, datos de audio, datos codificados, mensajes de correo elec-

trónico, anexos de correo electrónico, mensajes de conversación, otros tipos de mensajes de sistemas de mensajería, documentos transmitidos por facsímil y datos de interfaz del usuario y una combinación de los mismos.

18. El procedimiento de cualquiera de las reivindicaciones 1-17 en el que los datos se obtienen a partir de una pluralidad de fuentes de entrada, cada fuente de entrada proporcionando un tipo diferente de datos.

19. El procedimiento de cualquiera de las reivindicaciones 8-18 en el que la retroalimentación a partir del análisis de los datos se utiliza para determinar dicha por lo menos una característica de los datos.

20. Sistema para la gestión del almacenamiento de datos de acuerdo con el contenido de los datos obtenidos a partir de una o más fuentes de entrada (12, 36; 52):

- un módulo de formato (16) para convertir los datos a un formato común;

- un analizador (18; 38; 78) para analizar los datos en dicho formato común a fin de determinar automáticamente por lo menos una característica de los datos de acuerdo con el contenido;

- un motor de reglas (20; 40) para comparar dicha por lo menos una característica de los datos con por lo menos una regla; y

- un gestor de almacenamiento (22; 42) para recibir una decisión a partir de dicho motor de reglas y adaptarla, de acuerdo con dicha decisión TANTO seleccionando una de una pluralidad de medios de almacenamiento, dispositivo de almacenamiento, sistema de procedimiento o una combinación de los sismos y colocar los datos dentro o eliminar los datos de dicho medio de almacenamiento, sistema de almacenamiento, dispositivo de almacenamiento seleccionado o combinación de los mismos; COMO seleccionando no almacenar los datos.

21. El sistema de la reivindicación 20 adicionalmente comprendiendo una fuente de entrada (12; 36; 52) para proporcionar datos a dicho analizador en el que dicho motor de reglas determina si los datos se van a utilizar como retroalimentación a dicha fuente de entrada (12; 36; 52).

22. El sistema de cualquiera de las reivindicaciones 20 o 21 en el que el funcionamiento de dicho motor de reglas (20; 40) es accionado manualmente.

23. El sistema de cualquiera de las reivindicaciones 20 o 21 en el que el funcionamiento de dicho motor de reglas (20; 40) es accionado automáticamente.

24. El sistema de la reivindicación 23 en el que dicho motor de reglas (20; 40) es un iniciador de un proceso para por lo menos almacenar los datos.

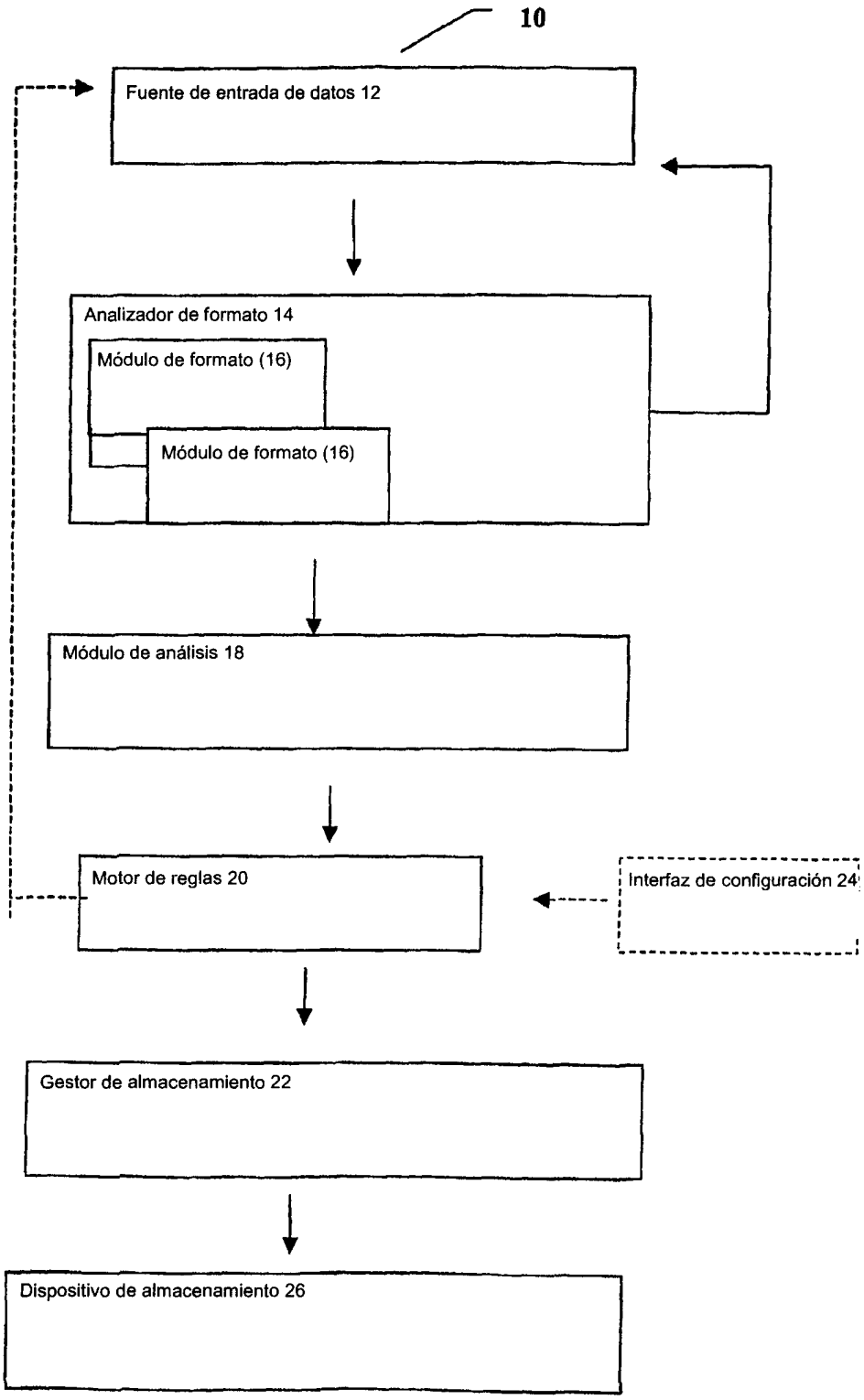
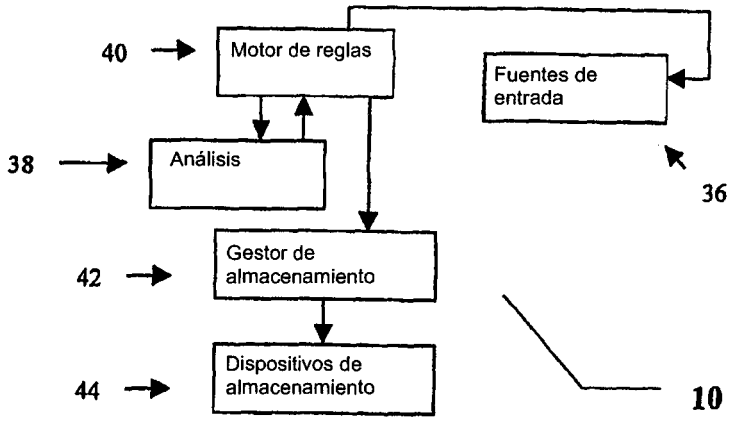

10
Fuente de entrada de datos 12
Analizador de formato 14
Módulo de formato (16)
Módulo de formato (16)
Módulo de análisis 18
Motor de reglas 20
Interfaz de configuración 24
Gestor de almacenamiento 22
Dispositivo de almacenamiento 26

Figura 1

Figura 2


40
Motor de reglas
Fuentes de entrada
38
Análisis
36
42
Gestor de almacenamiento
44
Dispositivos de almacenamiento
10

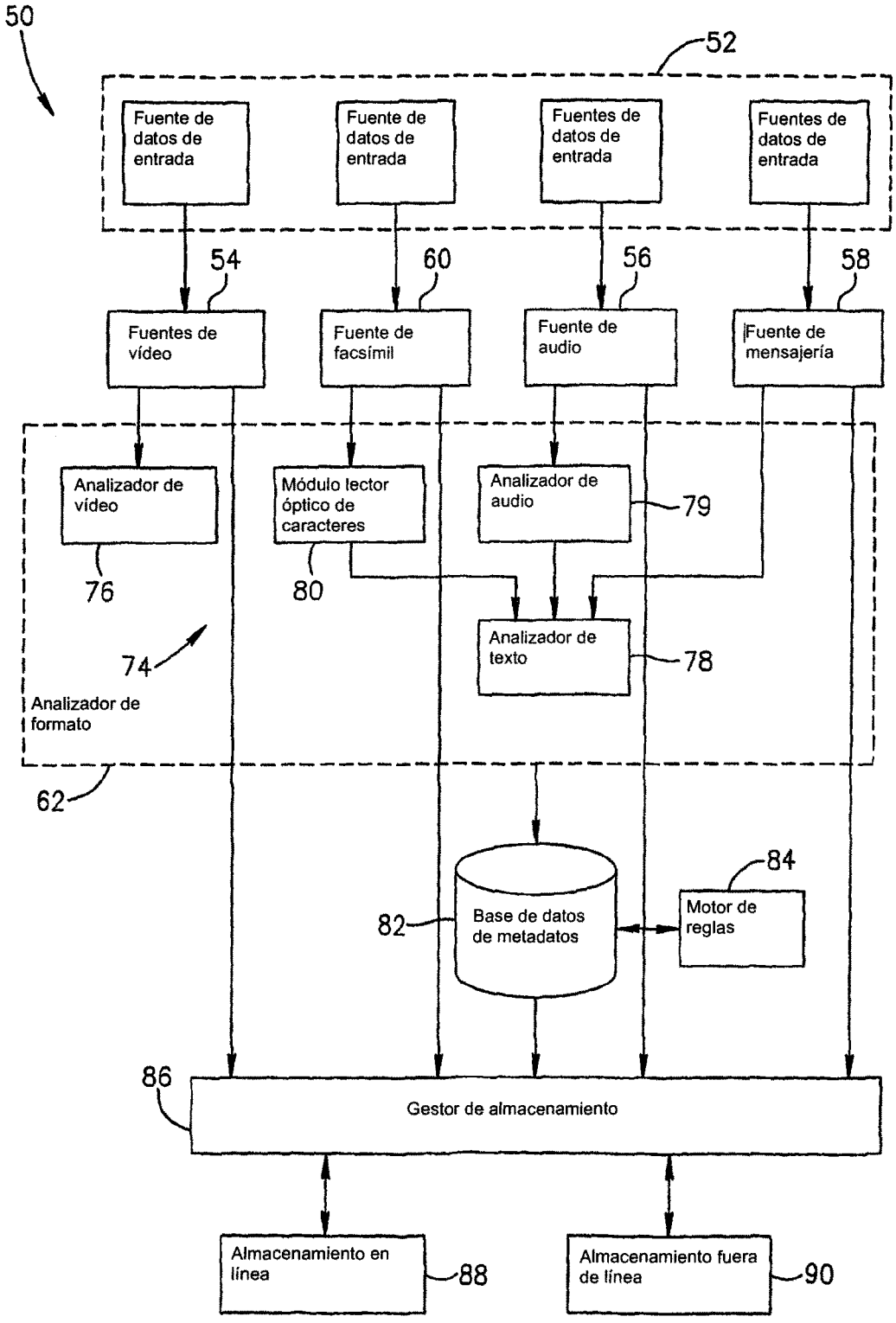

50
52
Fuente de datos de entrada
Fuente de datos de entrada
Fuentes de datos de entrada
Fuentes de datos de entrada
54
60
56
58
Fuentes de vídeo
Fuente de facsímil
Fuente de audio
Fuente de mensajería
Analizador de vídeo
Módulo lector óptico de caracteres
Analizador de audio
79
76
80
Analizador de texto
78
74
Analizador de formato
62
84
82
Base de datos de metadatos
Motor de reglas
86
Gestor de almacenamiento
Almacenamiento en línea
88
Almacenamiento fuera de línea
90

FIG.3