

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 734 058**

51 Int. Cl.:

**G06F 16/93** (2009.01)

**G06F 17/24** (2006.01)

**G06F 21/62** (2013.01)

**G06Q 50/18** (2012.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.12.2015** **E 15203167 (0)**

97 Fecha y número de publicación de la concesión europea: **08.05.2019** **EP 3188036**

54 Título: **Un método y un sistema para proporcionar un extracto de documento**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**04.12.2019**

73 Titular/es:  
**LEGALXTRACT APS (100.0%)**  
**Fredericiavej 16**  
**7100 Vejle, DK**

72 Inventor/es:  
**LAURSEN, RENÉ RICHARD**

74 Agente/Representante:  
**SÁEZ MAESO, Ana**

ES 2 734 058 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Un método y un sistema para proporcionar un extracto de documento

Campo de la invención

La invención se refiere a un método para proporcionar un extracto de documento a partir de un documento fuente.

- 5 Adicionalmente, la invención se refiere a un sistema para proporcionar un extracto de documento a partir de un documento fuente mediante el uso de dicho método.

Antecedentes de la invención

10 En Dinamarca, la Ley de Información Pública Danesa, que se aplica a la mayoría de las agencias públicas, oficinas administrativas públicas, etc. y que adicionalmente se extiende a ciertos proveedores privados y públicos de energía, etc., otorga a terceros, tales como periodistas, luego de solicitud el derecho para obtener acceso a ciertos documentos, archivos, etc. En otros países se aplican normas similares o correspondientes, tales como las leyes mencionadas como, por ejemplo, "Ley de Acceso a la Información Pública", "Ley de Libertad de Información", etc., que garantiza que el público, por ejemplo, un miembro del público, un periodista, etc. pueda tener acceso a archivos, documentos en dichos archivos, etc.

15 Sin embargo, en relación con dicho acceso a archivos, documentos, etc., en, por ejemplo, la administración pública, que puede haber sido otorgado a un tercero, se requiere que los documentos respectivos se examinen cuidadosamente para información, tal como por ejemplo nombres de ciertas personas, información clasificada, información confidencial, etc. que se debe evitar proporcionar al público en relación con los documentos respectivos.

20 Actualmente, esto se realiza en Dinamarca mediante un proceso manual que demanda tiempo y recursos relativos, por el cual el documento relevante se imprime en un papel, una persona legalmente calificada marca las palabras u otra información que se debe evitar este disponible al público, en el documento. Luego el documento con las marcas se presenta a una persona supervisora legalmente calificada para aprobación. En caso de aprobación, el documento en papel con las palabras marcadas u otra información marcada se reenvía a una persona legalmente calificada, que tacha manualmente las palabras marcadas con un rotulador negro. Posteriormente, el documento se escanea en un documento en formato pdf, que se imprime. De aquí en adelante, este "extracto" de documento resultante se examina para detectar si alguna de las palabras marcadas o la información marcada todavía son reconocibles y/o legibles, por ejemplo, si algunas de las letras son visibles a través de la marca negra. Si este es el caso, el tachado con el rotulador negro y el escaneo, impresión y examen posterior se repiten hasta obtener un resultado satisfactorio.

30 Se observa que, en las organizaciones administrativas danesas, etc., actualmente no está permitido utilizar programas informáticos disponibles, tales como programas de procesamiento de textos durante dicho proceso de extracción, ya que, por ejemplo, dichos programas generarán archivos temporales locales almacenados automáticamente, lo que pondrá en duda la seguridad de utilizar dichos programas. A este respecto, cabe notar que es un requisito que cuando un documento extraído resultante se reenvíe al tercero que haya solicitado el acceso, este tercero no podrá obtener ninguna información con respecto a las palabras u otra información que se haya tachado en el extracto de documento, sin importar si el tercero recibe el extracto de documento como un documento de papel o un documento electrónico.

40 Como quedará claro a partir de lo anterior, el trabajo y el tiempo involucrados en la producción de dichos extractos de documentos para el acceso público es considerable. A esto se puede agregar que, como consecuencia de las modificaciones introducidas en la versión más reciente de la Ley de Información Pública Danesa en vigor a partir del 1 de enero de 2014, que ha aumentado el número de solicitudes de acceso público permitidas, los recursos necesarios para su manejo se han incrementado aún más.

45 Cabe notar que actualmente los programas informáticos y los métodos asistidos por ordenador son conocidos en la técnica anterior para uso en relación con la realización de la redacción y/o sanitización de documentos que contienen, por ejemplo, información sensible. Aparentemente, el término para realizar la redacción se utiliza con frecuencia en relación con la eliminación de información sensible en un documento, por ejemplo, al tachar u ocultar, y el término para realizar la sanitización es con frecuencia como una generalización de la redacción, en la que los términos sensibles se pueden reemplazar por términos menos sensibles en lugar de tachar u ocultar los términos sensibles, por lo que la información útil aún se transmite al lector. Cabe notar que, los términos "redacción" y "sanitización" parecen utilizarse en diversos aspectos y significados dentro de este campo en particular. Sin embargo, como se mencionó anteriormente, dichos programas informáticos actuales y métodos asistidos por ordenador pueden poner en duda la seguridad, ya que, por ejemplo, dichos programas pueden generar archivos temporales locales almacenados automáticamente, etc., lo que puede generar el riesgo de que un tercero pueda obtener información con respecto a los términos sensibles eliminados.

55 El documento US 2005/0004922 A1 divulga un ejemplo de un programa informático para sanitización, que en este documento US se define como la eliminación de información sensible de un documento. El programa informático

5 comprende una función de escaneo y bases de datos para identificar información sensible tales como nombres y direcciones en un documento de fuente digital. La información sensible se muestra a un usuario con una lista de los términos de caso generales propuestos para sustitución. El usuario revisa las sustituciones propuestas y la lista revisada se guarda para uso al finalizar el documento de sustitución (y cualquier documento futuro). Los términos sensibles ya no se pueden ver en una pantalla que muestra el documento de sustitución, por ejemplo, en un programa de procesador de textos después de la finalización de la sustitución con la lista guardada que vincula los términos de caso sensibles y generales.

10 Por lo tanto, subsiste la necesidad de mejorar los métodos utilizados actualmente para reducir el tiempo y el esfuerzo utilizado para proporcionar dichos documentos a las personas que solicitaron y se les concedió acceso público, cuyos documentos se mencionaran como extractos de documentos, es decir, documentos en los que la información de carácter confidencial o información que, por otras razones, debe estar oscurecida "oculta".

Adicionalmente, subsiste la necesidad de proporcionar dicho proceso mejorado, que se pueda realizar utilizando un mayor grado de automatización, por ejemplo, mediante el uso de procesos asistidos por ordenador.

15 Aún más, subsiste la necesidad de dicho proceso mejorado, por medio del cual se pueda lograr un mayor grado de seguridad. También, es un objetivo lograr, por ejemplo, un mayor grado de aceptabilidad de los documentos extraídos en la primera versión producida, por lo que el tiempo y el esfuerzo involucrados se pueden reducir mientras se mantiene el nivel de calidad requerido, por ejemplo, el nivel de seguridad.

Adicionalmente, subsiste la necesidad de dicho proceso mejorado, por lo que se puede proporcionar un método flexible para, por ejemplo, el trabajo de oficina, las rutinas de trabajo, etc.

20 Estos y otros objetos son alcanzables por la invención como se explica con más detalle a continuación.

#### Resumen de la invención

La invención se refiere a un método para proporcionar un extracto de documento a partir de un documento fuente, dicho documento fuente es un documento clasificado, dicho método comprende las etapas de

- a) proporcionar dicho documento fuente en un formato legible por ordenador,
- 25 b) seleccionar por lo menos un elemento desde dicho documento fuente,
- c) establecer un conjunto de datos de identificación para identificar dicho por lo menos un elemento que se ha seleccionado,
- d) validar dicho por lo menos un elemento que se ha seleccionado,
- 30 e) proporcionar el extracto de documento en un formato fijo al realizar una conversión irreversible de dicho documento fuente, con base en dicho documento fuente y dicho conjunto de datos de identificación para dicho por lo menos un elemento que se ha validado, y

en el que el conjunto de datos de identificación por medio del cual dicho por lo menos se identifica un elemento que se ha seleccionado y/o validado, se almacena junto con una identificación de documento fuente.

35 De esta manera, se logra que se pueda proporcionar un extracto de documento por medio de un método asistido por ordenador y por lo cual el documento fuente permanece sin modificar, es decir debido a que los elementos seleccionados se identifican por un conjunto de datos de identificación, que se separa del documento fuente como tal. Adicionalmente, al proporcionar el extracto de documento a través de una conversión irreversible, no será posible desde el extracto de documento resultante recuperar cualquier información con respecto a los elementos seleccionados y validados.

40 Adicionalmente, se logra un método eficiente, por lo que el documento fuente no se modifica, es decir, debido a que los elementos seleccionados se identifican mediante un conjunto de datos de identificación, que está separado del documento fuente como tal, y por lo que además se facilita que el trabajo se puede interrumpir y reanudar más tarde, por ejemplo, al volver a cargar el documento fuente y el conjunto de datos de identificación almacenados por separado para los elementos ya seleccionados.

45 Por el término "documento clasificado" se entenderá para los propósitos de esta solicitud un documento que no se ha publicado anteriormente (y por lo tanto ya está disponible para cualquiera) y que potencialmente puede incluir información sensible, en la que el carácter de dicha información sensible puede estar muy extendida y puede incluir, por ejemplo, información de privacidad, información que se debe mantener en secreto, información relacionada con el secreto comercial, etc.

50 Por el término "formato fijo" se entenderá, para los propósitos de esta solicitud, un documento digital que tiene una imagen o diseño fijos. El documento no se puede editar para revelar información previa o histórica antes de la

conversión a un documento de formato fijo. Un documento en un formato fijo solo puede ser modificado al agregar nueva información al diseño o imagen original del documento cuando se convierte.

5 Ejemplos de documentos de formato fijo y programas informáticos para presentar documentos de "formato fijo" son el Formato de documento portátil (PDF) de Adobe Systems y la Especificación de Papel XML Abierto (OpenXPS) de Microsoft Corporation.

10 El conjunto o los conjuntos de datos de identificación para identificar uno o más de dichos por lo menos un elemento que se ha seleccionado se pueden establecer de varias maneras o formas, por ejemplo, un elemento se puede identificar por el número de página en el documento fuente y las coordenadas en la página, etc. El nombre del documento fuente también puede ser parte del conjunto o conjuntos de datos de identificación, por ejemplo, junto con el tamaño del documento fuente para garantizar una identificación segura del documento fuente correcto mediante comparación de tamaño.

En una realización de la invención, las etapas b) y c) se repiten para dicho documento fuente, antes de la etapa d) se realiza para el documento fuente en su totalidad.

De esta manera, se logra un método eficiente.

15 En una realización de la invención, la etapa d) para validar dicho por lo menos un elemento que se ha seleccionado comprende reconocer por lo menos un elemento seleccionado o rechazar por lo menos un elemento que se ha seleccionado.

De esta manera, se logra que se proporcione una posibilidad de realizar correcciones, si cualquiera, de los elementos seleccionados de una manera fácil de utilizar y eficiente en recursos.

20 En una realización de la invención la etapa b) y la etapa c) se repiten posteriores a la etapa d) y antes de la etapa e).

De esta manera, se proporciona un método flexible y fácil de utilizar.

25 En una realización de la invención, la etapa e) para proporcionar el extracto de documento al realizar una conversión irreversible de dicho documento fuente, con base en dicho documento fuente y dicho conjunto de datos de identificación para dicho por lo menos un elemento que se ha validado comprende enmascarar en el extracto de documento dicho por lo menos un elemento que se ha validado.

De esta manera, se logra que el extracto de documento corresponda al documento fuente con respecto a por ejemplo el formato, configuración, etc. y que es inmediatamente reconocible cuando los elementos se han hecho ininteligibles para el tercero.

30 En una realización de la invención, la conversión irreversible de acuerdo con la etapa e) comprende la conversión del documento fuente que está en una versión de extracto intermedio con por lo menos un elemento que se ha validado enmascarado en un documento de imagen, posiblemente seguido por una conversión en un formato de documento portátil.

De este modo, se consigue que la información sobre los elementos seleccionados y validados no se pueda recuperar del extracto de documento resultante.

35 El término "documento de imagen" se entenderá para los fines de esta aplicación como un documento digital definido por valores gráficos para mostrar una imagen en una pantalla de ordenador y para una copia impresa. El valor gráfico de una imagen solo revela la información gráfica y de posición necesaria, como el color de un píxel específico en la pantalla del ordenador (y en una copia impresa) para mostrar esta parte del documento de la imagen. Los valores gráficos de un documento de imagen no incluyen información o código que pueda ayudar a  
40 detectar un origen del documento de imagen, como los elementos seleccionados y validados mencionados anteriormente.

En una realización de la invención, el documento fuente se proporciona como un documento de texto.

De este modo, se logra que se puedan buscar elementos tales como palabras, nombres, abreviaturas, acrónimos, números, etc. utilizando, por ejemplo, el reconocimiento OCR.

45 El documento de texto puede comprender diferentes elementos que pueden ser objeto de extracción con la presente invención, tales como texto y/o elementos gráficos. Los elementos de texto pueden incluir palabras; nombres de personas, lugares y/o cosas; abreviaturas, acrónimos, números, etc. que se pueden buscar utilizando, por ejemplo, el reconocimiento OCR. Los elementos gráficos pueden incluir fotografías, dibujos u otras imágenes visuales; símbolos; representaciones gráficas; elementos de texto que no han sido escaneados OCR, etc.

50 El formato digital de un documento de texto como se definió anteriormente puede ser cualquier formato que se utilice generalmente para trabajar con documentos utilizando medios informáticos, por ejemplo, formatos de programas de procesador de textos tales como Microsoft Word (archivos .doc), formatos de programas de formato fijo tales como

Adobe Acrobat (archivos .pdf), formatos de programas de dibujo como Autodesk Autocad (archivos .dwg), formatos de documentos relacionados con Internet (archivos .xml o similares), etc., que pueden ser objeto de extracción con la presente invención.

5 El documento fuente en un formato de un documento de texto se puede cargar en el aparato informático desde, por ejemplo, un archivo electrónico o el documento puede escanearse y cargarse en el aparato informático. También se pueden utilizar otras formas de proporcionar y cargar el documento fuente.

En una realización de la invención, por lo menos un elemento que se ha seleccionado de dicho documento fuente puede ser uno de

- una palabra,

10 - una pluralidad de palabras en secuencia,

- un párrafo,

- una casilla y

- combinaciones de los anteriores.

15 En una realización de la invención, la casilla puede comprender una foto, una imagen, un dibujo, un diagrama y/o una palabra.

En una realización de la invención, la etapa b) de seleccionar por lo menos un elemento desde dicho documento fuente se facilita por uno de

- utilizar una funcionalidad de enfoque que utiliza por ejemplo reconocimiento OCR,

- marcar una pluralidad de palabras, un párrafo y/o un área de documento.

20 De esta manera, se proporciona un método flexible y fácil de utilizar, que adicionalmente facilita un sistema eficiente en costes y tiempo para proporcionar extractos de documentos.

En un segundo aspecto de la invención, se proporciona un sistema para proporcionar un extracto de documento a partir de un documento fuente utilizando un método de acuerdo con una cualquiera de las reivindicaciones 1

25 - 10, dicho sistema comprende un aparato ordenador, medios de visualización y medios de entrada, dicho sistema se configura para

- visualizar dicho documento fuente sobre dichos medios de visualización,

- facilitar por lo menos un elemento desde dicho documento fuente que se va a seleccionar de una forma sin modificar el documento fuente,

30 - establecer un conjunto de datos de identificación para identificar dicho por lo menos un elemento que se ha seleccionado,

- facilitar un proceso de validación de dicho por lo menos un elemento que se ha seleccionado,

- y proporcionar el extracto de documento en un formato fijo luego de un proceso de validación completo al realizar una conversión irreversible de dicho documento fuente, con base en dicho documento fuente y dicho conjunto de datos de identificación para dicho por lo menos un elemento que se ha validado,

35 en el que el sistema adicionalmente se configura para almacenar dicho conjunto de datos de identificación por medio del cual dicho por lo menos se identifica un elemento que se ha seleccionado y/o validado, junto con una identificación de documento fuente.

40 De esta manera, se logra que se pueda proporcionar un extracto de documento por medio de un aparato ordenador y por lo que el documento fuente permanece sin modificar, es decir debido a los elementos seleccionados que se identifican mediante un conjunto de datos de identificación, que está separado del documento fuente como tal. Adicionalmente, al proporcionar el extracto de documento a través de una conversión irreversible, no será posible, a partir del extracto de documento resultante, recuperar alguna información sobre los elementos seleccionados y validados.

45 Adicionalmente, se logra un sistema eficiente y fácil de utilizar, por lo que el documento fuente no se modifica, es decir, debido a que los elementos seleccionados se identifican mediante un conjunto de datos de identificación, que está separado del documento fuente como tal, y por lo que además se facilita que el trabajo se pueda interrumpir y reanudar más tarde, por ejemplo, al recargar el documento fuente y el conjunto de datos de identificación almacenados por separado para los elementos ya seleccionados.

5 El experto en la técnica entenderá que el aparato informático comprende medios procesadores, por ejemplo, medios procesadores para facilitar la visualización del documento fuente y otros documentos sobre el medio de visualización, para ejecutar etapas operativas de programas informáticos, por ejemplo, etapas de un programa de aplicación de acuerdo con una realización de la invención, para operar el aparato informático de acuerdo con la entrada procedente de medios de entrada tales como el ratón, el teclado, etc. También, se entenderá que el aparato informático comprende medios de almacenamiento, por ejemplo, medios de almacenamiento para uso como se ejemplifica en la siguiente descripción detallada. También, el aparato informático puede comprender y/o estar conectado a otros dispositivos y/o elementos normalmente utilizados, tales como lectores de medios legibles por ordenador. También se observa que el aparato informático puede ser parte de una red informática, por ejemplo, una red de área amplia (WAN) o local (LAN) o posiblemente a través de Internet. Cuando el aparato informático es parte de una red, el programa de aplicación puede, por ejemplo, ejecutarse por lo menos parcialmente en un ordenador remoto o el aparato informático puede ser un ordenador autónomo. También resultará evidente para un experto en la materia que el aparato informático y la red informática, en caso de que el aparato informático forme parte de dicha red informática, contará con medidas de protección de vanguardia tales como cortafuegos, software informático anti-hackear, etc.

En una realización de la invención, el sistema se puede configurar para facilitar la selección de por lo menos un elemento desde dicho documento fuente por uno de

- utilizar una funcionalidad de enfoque que utiliza por ejemplo reconocimiento OCR, y
- marcar una pluralidad de palabras, un párrafo y/o un área de documento.

20 De esta manera, se proporciona un sistema flexible y fácil de utilizar, que adicionalmente facilita un método eficiente en costes y tiempo para proporcionar extractos de documentos.

En una realización de la invención, el sistema se puede configurar para realizar dicha conversión irreversible mediante una conversión del documento fuente está en una versión de extracto intermedio con por lo menos un elemento que se ha validado enmascarado en un documento de imagen, posiblemente seguido por una conversión en un formato de documento portátil.

De esta manera, se logra que la información acerca de los elementos seleccionados y validados no se pueden recuperar del extracto de documento resultante.

30 En un tercer aspecto de la invención, se proporciona un producto de programa informático, dicho producto de programa informático que comprende instrucciones legibles por ordenador para llevar a cabo todas las etapas del método de una cualquiera de las reivindicaciones 1-10, cuando el producto de programa informático se ejecuta sobre un sistema informático adecuado.

35 En lo que antecede, se han descrito el método y el sistema para uso en relación con las Leyes de Información Pública o similares, en las que los documentos extraídos se proporcionan en respuesta a solicitudes de acceso concedidas a, por ejemplo, documentos administrativos públicos, archivos, etc. Sin embargo, la invención también se puede utilizar en otros campos y aplicaciones.

#### Las Figuras

La invención se explicará con más detalle a continuación con referencia a las Figuras de las cuales

la Figura 1 muestra un ejemplo de un flujo de trabajo de acuerdo con una realización de la invención,

La Figura 2 muestra un ejemplo adicional de un flujo de trabajo de acuerdo con una realización de la invención,

40 La Figura 3 ilustra un ejemplo de una interfaz gráfica de usuario para un programa de aplicación de extracción de acuerdo con una realización de la invención, y

La Figura 4 ilustra realizaciones de ejemplo adicionales de acuerdo con la invención.

#### Descripción detallada

45 En la Figura 1 se muestra un ejemplo de un flujo de trabajo de acuerdo con una realización de la invención. De acuerdo con este ejemplo de flujo de trabajo, se activa un programa de aplicación de extracción y desde este programa de aplicación se carga un documento fuente (en 1) en un aparato de ordenador o dispositivo de ordenador adecuado, por ejemplo, un ordenador portátil, un ordenador estacionario, etc., y se muestra al usuario en el medio de pantalla correspondiente. El documento fuente puede ser un documento que se reenviará a una persona que haya solicitado acceso a un archivo, en el que está contenido el documento fuente. El documento fuente, que puede estar en un formato de texto, se puede cargar en el aparato informático desde, por ejemplo, un archivo electrónico o el documento puede escanearse y cargarse en el aparato informático. También se pueden utilizar otras formas de proporcionar y cargar el documento fuente.

- 5 Cuando el documento fuente se ha cargado y se muestra en los medios de visualización, el usuario puede buscar (en 2) el documento en busca de ciertas palabras, nombres, abreviaturas, acrónimos, números, etc., por ejemplo, al utilizar un método OCR para detectar ciertas palabras. La búsqueda se puede iniciar utilizando medios de entrada tales como el teclado, el mouse del ordenador u otros medios de entrada del ordenador. Adicionalmente, una o más de las palabras reconocidas por OCR se pueden enfocar al navegar hacia la palabra utilizando el teclado o el mouse del ordenador. Cuando el programa de aplicación enfoca una palabra reconocida por OCR, la palabra se marcará utilizando, por ejemplo, una primera marca de color, mejora o similar para indicar que la palabra es una palabra reconocida por OCR.
- 10 Las palabras enfocadas (en 3) se pueden revisar y seleccionar posteriormente, lo que se indica mediante una marca que utiliza, por ejemplo, un segundo color de marca, mejora o similar que es diferente de la primera marca para indicar que el usuario ha seleccionado una o mas palabras
- Adicionalmente, cuando se seleccionan dos o más palabras reconocidas por OCR, que se colocan una al lado de la otra, las palabras, así como el espacio entre las palabras, se marcan como una marca continua.
- 15 Adicionalmente, se proporcionan otras formas de seleccionar elementos del documento fuente como se indica en 4. Por ejemplo, en un modo de párrafo, se puede seleccionar una pluralidad de palabras reconocidas por OCR mediante, por ejemplo, el ratón del ordenador, por medio del cual se puede seleccionar una casilla. Se puede definir, al cubrir la pluralidad de palabras en, por ejemplo, un párrafo. De acuerdo con otro ejemplo, se pueden seleccionar otras palabras reconocidas por OCR en un modo de casilla, mediante la cual se puede definir una casilla, por ejemplo, el mouse del ordenador, que puede cubrir elementos como imágenes, dibujos, diagramas, palabras que no han sido reconocidos por OCR, etc.
- 20 Como se explicará con más detalle a continuación en relación con la Figura 2, las marcas de los elementos seleccionados en el documento se pueden guardar utilizando una funcionalidad de guardar. El documento fuente permanece sin modificaciones, pero los datos para identificar los elementos marcados se guardan en un archivo intermedio o temporal junto con una identificación del documento fuente. Cuando se reanuda el trabajo, el documento fuente respectivo se vuelve a cargar junto con el archivo intermedio o temporal que contiene los datos para identificar los elementos marcados.
- 25 Volviendo a la Figura 1, el programa de aplicación proporciona una función de validación en 5, en la que, por ejemplo, un supervisor o similar puede revisar los elementos seleccionados y, por lo tanto, marcados en el documento.
- 30 Al finalizar la validación en 5, el extracto de documento resultante puede generarse en 6 en el sentido de que los elementos seleccionados y validados están enmascarados, por ejemplo, completamente cubiertos, reemplazados o similares con color negro, por ejemplo, con una casilla negra, para evitar completamente que cualquiera de los elementos sea reconocible y se realice una conversión irreversible, por ejemplo, en un documento de imagen para evitar que cualquier información sobre los elementos seleccionados, validados y enmascarados se pueda recuperar del extracto de documento resultante.
- 35 Posteriormente, el extracto de documento resultante en formato de imagen en 7 se puede convertir en un formato de documento portátil (pdf) para facilitar el manejo y envío del extracto de documento resultante a la persona o tercero que haya solicitado el acceso al documento.
- 40 En la Figura 2 se muestra un flujo de trabajo esencialmente como se discutió en relación con la Figura 1, pero adicionalmente se ejemplifica aquí en relación con la búsqueda 2, revisión y selección 3, 4 es posible que el usuario salte libremente entre las distintas etapas indicadas por los bucles 9 de retorno.
- También, se muestra en la Figura 2 que en relación con la función 5 de validación, en la que, por ejemplo, un supervisor o similar puede revisar los elementos seleccionados, y por lo tanto marcados, en el documento, es posible que el supervisor apruebe ("sí") o desapruebe ("no") los elementos seleccionados en el documento, en este último caso, la persona que ha realizado el trabajo puede enmendar o corregir, es decir, como lo indica el bucle 10 de retorno puntuado que permite al usuario volver a una etapa anterior.
- 45 Adicionalmente, se muestra una funcionalidad 8 de guardar, por lo que, en relación con cada etapa, se puede guardar el trabajo ya realizado, por ejemplo, las marcas de los elementos seleccionados en el documento se pueden guardar utilizando esta función de guardar. Mediante esta funcionalidad de guardar, el documento fuente permanece sin modificaciones, pero, por ejemplo, los datos para identificar los elementos marcados se guardan en un archivo intermedio o temporal junto con una identificación del documento fuente. Cuando se reanuda el trabajo, el documento fuente respectivo se vuelve a cargar junto con el archivo intermedio o temporal que contiene los datos para identificar los elementos marcados. El trabajo se puede reanudar en la misma etapa en la que se guardó, pero en esencia se puede reanudar en cualquiera de las etapas 2, 3 y 4.
- 50 Como se indica, también puede ser posible que el supervisor, en relación con la función 5 de validación, utilizar la funcionalidad 8 de guardar como se indica mediante líneas puntuadas.
- 55

La Figura 3 ilustra un ejemplo de una interfaz gráfica de usuario para un programa de aplicación de extracción de acuerdo con una realización de la invención, en el que se muestran un editor 20 y un visor 40.

5 El editor comprende, por ejemplo, una clave 22 para abrir un documento fuente, por ejemplo, para encontrar y cargar el documento, una clave 24 para guardar el trabajo realizado, por ejemplo, guardar juntos los datos relacionados con el trabajo en un archivo intermedio o temporal junto con una identificación única del documento fuente, una clave 26 para seleccionar un elemento en el documento fuente y una clave 28 para realizar una extracción en el documento.

10 El usuario iniciará el trabajo en el editor 20 encontrando, cargando y abriendo el documento fuente respectivo, que en la Figura 3 se muestra como un ejemplo 32a relativamente sencillo. El usuario puede continuar posteriormente buscando elementos como palabras, seleccionando uno o más de estos y/o seleccionando otros elementos marcándolos con casillas como lo indica el documento fuente en la versión 32b seleccionada.

Después de haberse realizado una validación y al operar la tecla 28 de extracción, el extracto 42 del documento se mostrará en el visor 40 con los elementos seleccionados y validados respectivos ennegrecidos con casillas 44 negras.

15 La Figura 4 ilustra otras realizaciones de ejemplo del método y el sistema de acuerdo con la invención. En este documento, se muestra que en relación con la etapa a) de proporcionar un documento fuente en un formato legible por ordenador, por ejemplo, en formato pdf, el documento fuente, por ejemplo, se busca y carga 50 desde una fuente como una base de datos DB1.

20 Posteriormente, el trabajo relacionado con la búsqueda y selección 52 de elementos en el documento fuente y la etapa c) de establecer un conjunto de datos de identificación para identificar uno o más elementos que han sido seleccionados 54 implica una base de datos DB2, por ejemplo, una base de datos en relación con el programa de aplicación de extracción, en el que la base de datos que identifica el conjunto de datos por medio de los cuales dicho uno o más elementos que se ha/han seleccionado, se almacenan junto con una identificación del documento fuente. El conjunto de datos de identificación puede establecerse de varias maneras o formas, por ejemplo, un elemento puede identificarse por un número de página en el documento fuente y las coordenadas en la página, etc. El nombre del documento fuente también puede ser parte del conjunto o conjuntos de datos de identificación, por ejemplo, junto con el tamaño del documento fuente para garantizar una identificación segura del documento fuente correcto mediante comparación de tamaño.

30 Por lo tanto, el documento fuente permanece sin modificaciones, es decir, debido a que los elementos seleccionados se identifican mediante un conjunto de datos de identificación, que está separado del documento fuente tal como se proporciona y almacena en la base de datos DB1. Adicionalmente, de esta manera es posible que el trabajo se pueda interrumpir y reanudar más tarde, por ejemplo, al volver a cargar el documento fuente desde DB1 y el conjunto de datos de identificación almacenados por separado para los elementos ya seleccionados de DB2.

35 Finalmente, se muestra en la Figura 4 que la etapa d) de validar los elementos seleccionados en 56 y la etapa e) de realizar la extracción en el documento en 58 se realiza en interacción con otra base de datos DB3, por ejemplo, una base de datos relacionada con el programa de aplicación de extracción, en el que se almacena el extracto de documento.

40 El extracto de documento se puede renombrar automáticamente cuando se almacena en una base de datos, por ejemplo, DB3. El cambio de nombre se puede realizar, por ejemplo, al agregar una letra al nombre del documento fuente como "X-name.pdf" o al cambiar el nombre del documento fuente completamente, por ejemplo, con un generador de nombre de archivo. Una persona que realiza la extracción del documento fuente también puede cambiar manualmente el nombre del extracto de documento al almacenarlo en una base de datos.

Las bases de datos DB2 y DB3 se pueden ubicar sobre dispositivos de almacenamiento de datos separados en el mismo lugar o en diferentes lugares con enlaces de datos entre los dispositivos o se pueden ubicar en un dispositivo de almacenamiento de datos en diferentes áreas de almacenamiento del dispositivo.

45 En la descripción anterior, se han descrito diversas realizaciones de la invención con referencia a los dibujos, pero es evidente para un experto en la técnica que la invención se puede llevar a cabo en un número infinito de formas, utilizando, por ejemplo, los ejemplos divulgados en la descripción en varias combinaciones, y dentro de un amplio rango de variaciones dentro del alcance de las reivindicaciones adjuntas.

Lista de números de referencia

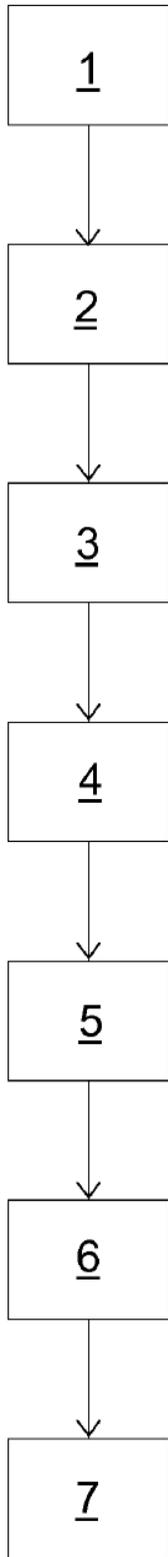
- 50 1 Se carga documento fuente  
 2 Búsqueda y enfoque  
 3 Revisión y selección  
 4 Otras formas de selección

- 5 Validación
- 6 Generación de extracto de documento por conversión irreversible
- 7 Conversión en un formato de documento portátil
- 8 Funcionalidad de guardar
- 5 9 Bucle de retorno
- 10 Bucle de retorno de la etapa de validación
- 20 Editor en el programa de aplicación de extracción
- 22 Clave para abrir un documento fuente
- 24 Clave para guardar el trabajo realizado
- 10 26 Clave para seleccionar un elemento
- 28 Clave para realizar una extracción en el documento
- 32a Documento fuente
- 32b Documento fuente en la versión seleccionada
- 40 Visor en el programa de aplicación de extracción
- 15 42 Extracto del documento mostrado en el visor
- 44 Elementos seleccionados y validados enmascarados/reemplazados por casillas negras
- 50 Suministro del documento fuente - etapa a)
- 52 Selección de elementos en el documento - etapa b)
- 54 Establecimiento de conjunto de datos para identificar elementos seleccionados - etapa c)
- 20 56 Validación de elementos seleccionados - etapa d)
- 58 Realización de extracción en el documento - etapa e)

**REIVINDICACIONES**

1. Un método para proporcionar un extracto de documento a partir de un documento fuente, dicho documento fuente es un documento clasificado, dicho método comprende las etapas de
- a) proporcionar (50) dicho documento (32a) fuente en un formato legible por ordenador,
  - 5 b) seleccionar (3; 52) por lo menos un elemento desde dicho documento (32a) fuente,
  - c) establecer (54) un conjunto de datos de identificación para identificar dicho por lo menos un elemento que se ha seleccionado,
  - d) validar (5; 56) dicho por lo menos un elemento que se ha seleccionado, caracterizado porque el método comprende
  - 10 e) proporcionar (6; 58) el extracto (42) del documento en un formato fijo al realizar una conversión irreversible de dicho documento fuente, con base en dicho documento (32a) fuente y dicho conjunto de datos de identificación para dicho por lo menos un elemento que se ha validado, y
- en el que dicho conjunto de datos de identificación por medio del cual dicho por lo menos se identifica un elemento que se ha seleccionado y/o validado, se almacena junto con una identificación de documento fuente.
- 15 2. El método de acuerdo con la reivindicación 1, en el que se repiten las etapas b) y c) (9) para dicho documento (32a) fuente, antes de la etapa d) se realiza para el documento fuente en su totalidad.
3. El método de acuerdo con la reivindicación 1 o 2, en el que la etapa d) para validar (5; 56) dicho por lo menos un elemento que se ha seleccionado comprende reconocer por lo menos un elemento seleccionado o rechazar por lo menos un elemento que se ha seleccionado.
- 20 4. El método de acuerdo con la reivindicación 3, en el que la etapa b) y la etapa c) se repiten (9) posterior a la etapa d) y antes de la etapa e).
5. El método de acuerdo con cualquiera de las reivindicaciones 1 - 4, en el que la etapa e) para proporcionar (6; 58) el extracto (42) del documento al realizar una conversión irreversible de dicho documento (32a) fuente, con base en dicho documento fuente y dicho conjunto de datos de identificación para dicho por lo menos un elemento que se ha validado comprende enmascarar en el extracto de documento dicho por lo menos un elemento que se ha validado.
- 25 6. El método de acuerdo con cualquiera de las reivindicaciones 1 - 5, en el que dicha conversión irreversible de acuerdo con la etapa e) comprende la conversión del documento fuente que está en una versión de extracto intermedio con por lo menos un elemento que se ha validado enmascarado en un documento de imagen, posiblemente seguido por una conversión en un formato de documento portátil.
- 30 7. El método de acuerdo con cualquiera de las reivindicaciones 1 - 6, en el que dicho documento (32a) fuente se proporciona como un documento de texto.
8. El método de acuerdo con cualquiera de las reivindicaciones 1 - 7, en el que dicho por lo menos un elemento que se ha seleccionado de dicho documento (32a) fuente puede ser uno de
- una palabra,
  - 35 - una pluralidad de palabras en secuencia,
  - un párrafo,
  - una casilla y
  - combinaciones de los anteriores.
- 40 9. El método de acuerdo con la reivindicación 8, en el que dicha casilla puede comprender una foto, una imagen, un dibujo, un diagrama y/o una palabra.
10. El método de acuerdo con cualquiera de las reivindicaciones 1 - 9, en el que dicha etapa b) de seleccionar (3; 52) por lo menos un elemento desde dicho documento (32a) fuente se facilita por uno de
- utilizar una funcionalidad de enfoque que utiliza por ejemplo reconocimiento OCR,
  - marcar una pluralidad de palabras, un párrafo y/o un área de documento.

11. Un sistema para proporcionar un extracto de documento a partir de un documento fuente utilizando un método de acuerdo con una cualquiera de las reivindicaciones 1 - 10, dicho sistema comprende un aparato ordenador, medios (20) de visualización y medios (22, 24, 26, 28) de entrada, dicho sistema se configura para
- visualizar dicho documento (32a) fuente sobre dichos medios (20) de visualización,
- 5 - facilitar por lo menos un elemento desde dicho documento fuente que se va a seleccionar (52) de una forma sin modificar el documento fuente,
- establecer (54) un conjunto de datos de identificación para identificar dicho por lo menos un elemento que se ha seleccionado,
- 10 - facilitar un proceso de validación (5; 56) de dicho por lo menos un elemento que se ha seleccionado, adicionalmente caracterizado porque el sistema se configura para proporcionar (6; 58) el extracto (42) del documento luego de un proceso de validación completo (5; 56) al realizar una conversión irreversible de dicho documento (32a) fuente, con base en dicho documento fuente y dicho conjunto de datos de identificación para dicho por lo menos un elemento que se ha validado,
- en el que dicho sistema se configura adicionalmente para almacenar (DB2) dicho conjunto de datos de identificación por medio del cual dicho por lo menos se identifica un elemento que se ha seleccionado y/o validado, junto con una identificación de documento fuente.
- 15
12. El sistema de acuerdo con la reivindicación 11, en el que dicho sistema se configura para facilitar la selección (3; 52) de por lo menos un elemento desde dicho documento (32a) fuente por uno de
- utilizar una funcionalidad de enfoque que utiliza por ejemplo reconocimiento OCR, y
- 20 - marcar una pluralidad de palabras, un párrafo y/o un área de documento.
13. El sistema de acuerdo con una cualquiera de las reivindicaciones 11 - 12, en el que dicho sistema se configura para realizar dicha conversión irreversible mediante una conversión del documento fuente está en una versión de extracto intermedio con por lo menos un elemento que se ha validado enmascarado en un documento de imagen, posiblemente seguido por una conversión en un formato de documento portátil.
- 25
14. Un producto de programa informático que comprende instrucciones legibles por ordenador para llevar a cabo todas las etapas del método de una cualquiera de las reivindicaciones 1-10, cuando el producto de programa informático se ejecuta sobre un sistema informático adecuado.



*Fig. 1*

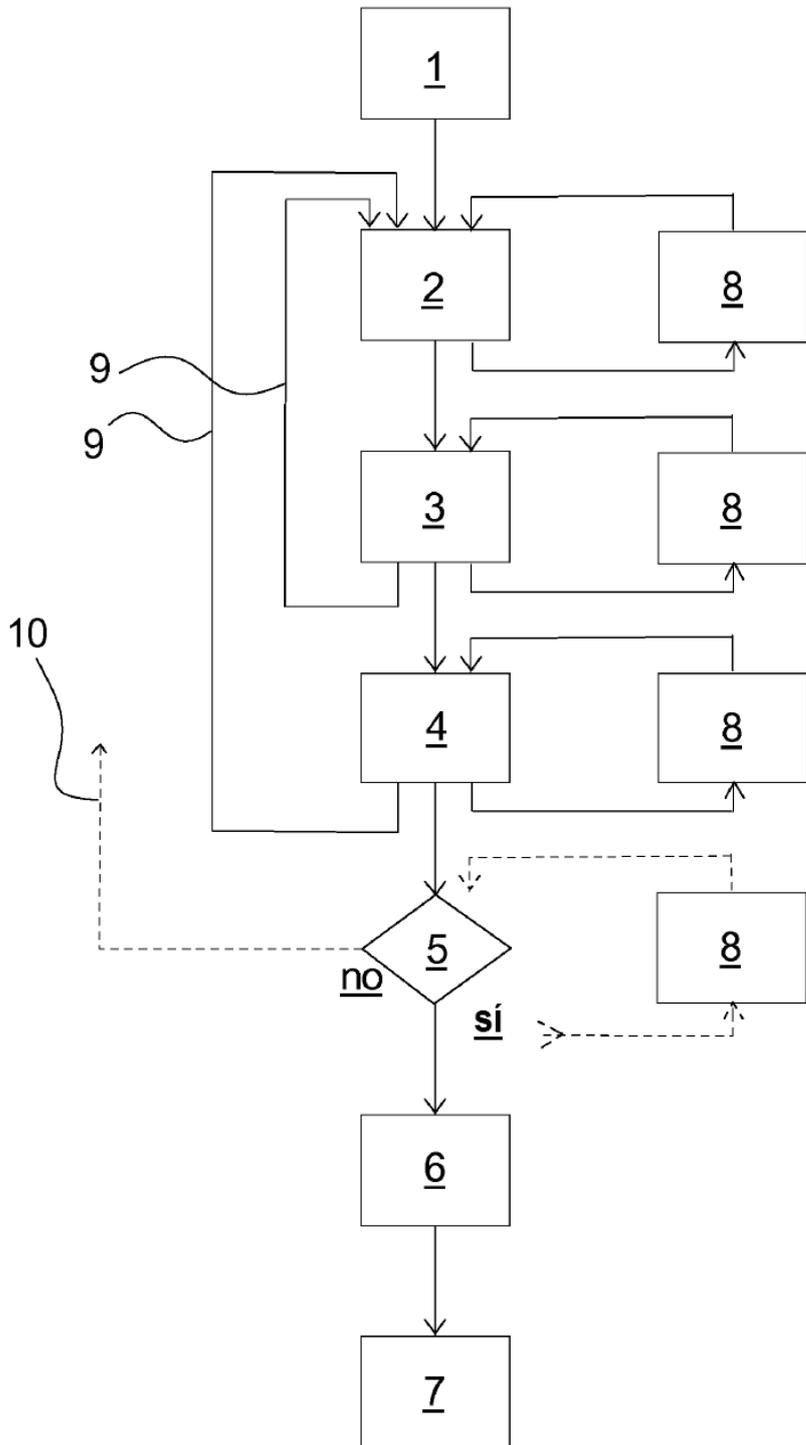


Fig. 2

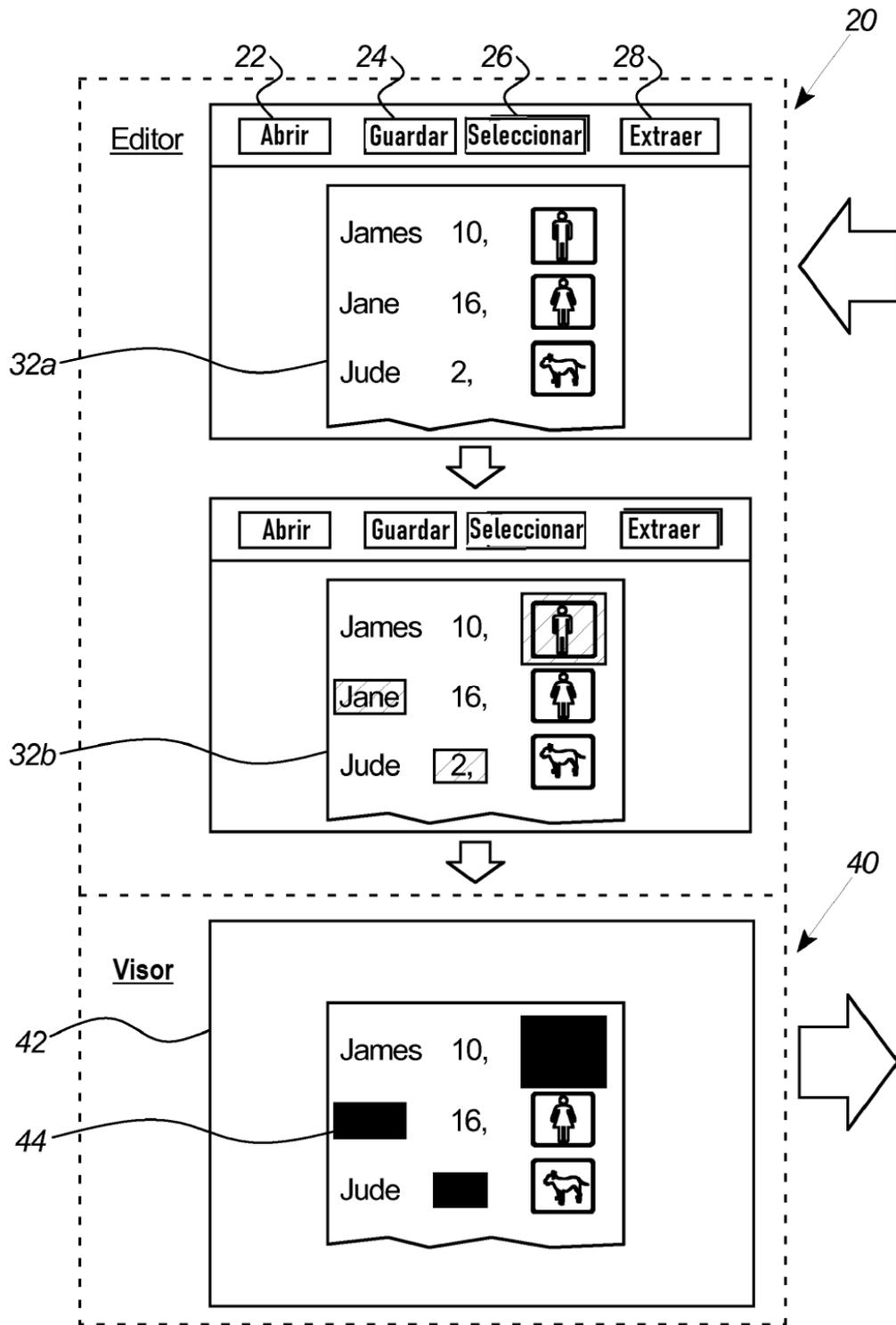


Fig. 3

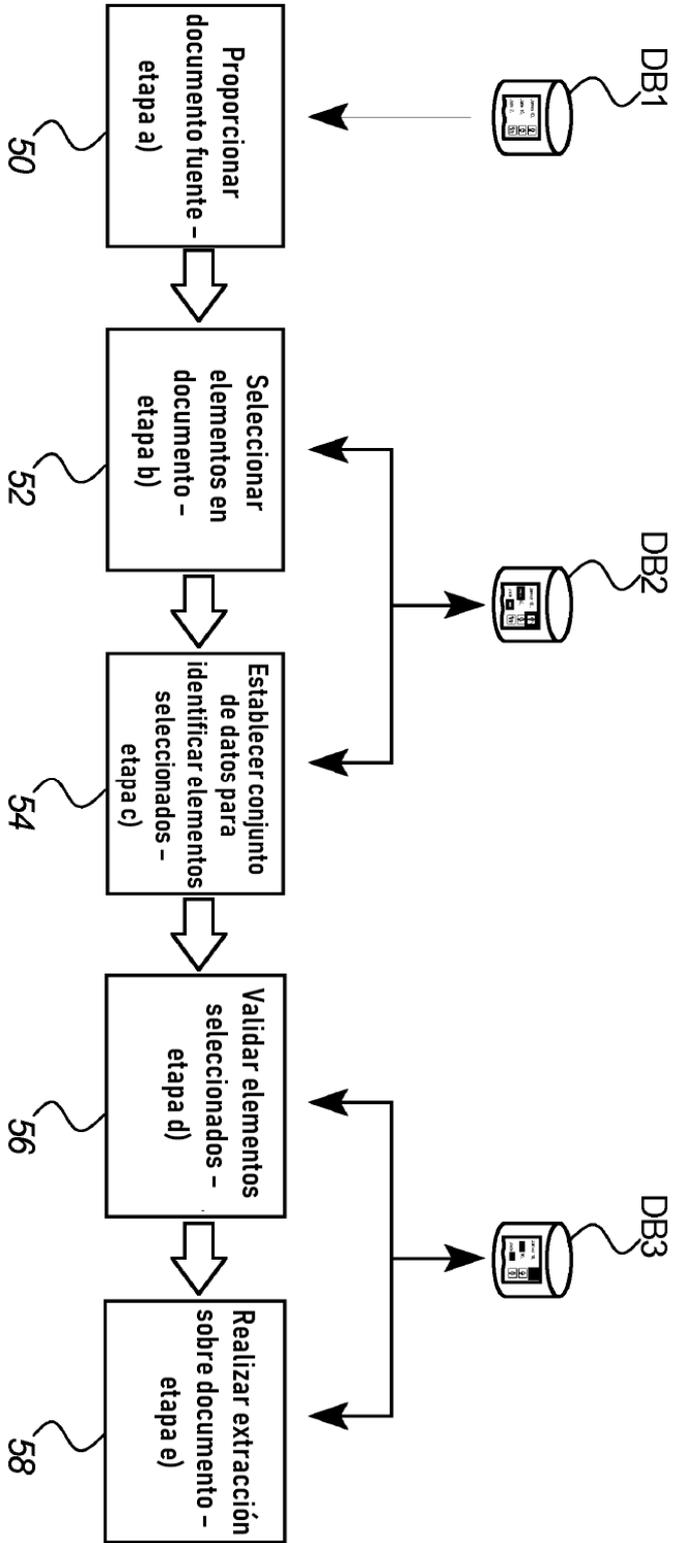


Fig. 4