

19



OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 555 180**

21 Número de solicitud: 201590093

51 Int. Cl.:

G06F 17/21 (2006.01)

G06F 17/20 (2006.01)

G06F 3/048 (2013.01)

12

SOLICITUD DE PATENTE

A2

22 Fecha de presentación:

27.02.2014

30 Prioridad:

28.02.2013 US 13/781446

43 Fecha de publicación de la solicitud:

29.12.2015

71 Solicitantes:

**THOMSON REUTERS GLOBAL RESOURCES (TRGR) (100.0%)
Neuhofstrasse 1
6340 Baar CH**

72 Inventor/es:

COOPER, Aaron

74 Agente/Representante:

CURELL AGUILÁ, Mireia

54 Título: **Método implementado por ordenador para sincronizar anotaciones entre un documento impreso y un documento electrónico, soporte legible por ordenador y sistema correspondientes**

57 Resumen:

Método implementado por ordenador para sincronizar anotaciones entre un documento impreso y un documento electrónico, soporte legible por ordenador y sistema correspondientes. Está prevista una imagen de una parte de documento impreso para un sincronizador. El sincronizador recupera una versión electrónica del documento impreso e identifica una parte de texto electrónico que es textualmente similar a una parte de texto impreso. El sincronizador detecta una anotación en la parte de documento impreso e inserta una anotación digital correspondiente en el documento electrónico.

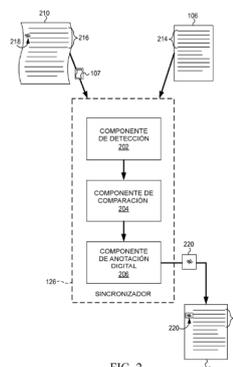


FIG. 2

DESCRIPCIÓN

Método implementado por ordenador para sincronizar anotaciones entre un documento impreso y un documento electrónico, soporte legible por ordenador y sistema correspondientes.

5

Referencia cruzada a solicitudes relacionadas

La presente solicitud reivindica los derechos de la solicitud no provisional US n.º 13/781.446, presentada el 28 de febrero de 2013, e incorpora expresamente la misma en su totalidad como referencia.

10

Antecedentes

La popularidad de los libros electrónicos y otros documentos electrónicos ha aumentado entre lectores, tales como profesionales jurídicos, profesionales de la medicina, estudiantes y otros, en los últimos años. Algunos de estos lectores mantienen una versión impresa de un documento, así como una versión electrónica del mismo. Muchos lectores se han acostumbrado a realizar anotaciones en documentos impresos resaltando y subrayando partes de texto, escribiendo notas en los márgenes, escribiendo notas entre líneas de texto impreso, tachando texto impreso, y similares. Aunque los libros electrónicos convencionales y otros documentos electrónicos permiten en ocasiones añadir directamente anotaciones digitales por medio de un dispositivo de lectura, el lector normalmente debe continuar remitiéndose a la versión impresa del documento en la que ha efectuado las anotaciones inicialmente, ya que estas plataformas de documentos electrónicos no posibilitan la sincronización de anotaciones entre un documento impreso y una versión electrónica del documento impreso.

15

20

25

Sumario

Formas de realización de la presente invención dadas a conocer en la presente facilitan la sincronización de anotaciones entre un documento impreso y un documento electrónico. En formas de realización, se proporciona a un sincronizador una imagen de una parte de un documento impreso. El sincronizador puede recuperar una versión electrónica del documento impreso e identificar una parte de texto electrónico que sea textualmente similar a una parte de texto impreso. Según formas de realización, el sincronizador detecta una anotación en la parte del documento impreso e inserta una anotación digital correspondiente en la parte similar del documento electrónico identificada textualmente.

30

35

Breve descripción de los dibujos

La figura 1 es un diagrama de bloques que ilustra un entorno de funcionamiento (y, en algunas formas de realización, aspectos de la presente invención) de acuerdo con formas de realización de la presente invención;

la figura 2 es un diagrama esquemático que representa un funcionamiento ilustrativo de un sincronizador de acuerdo con formas de realización de la presente invención;

la figura 3 es un diagrama esquemático que representa una operación de poda ilustrativa de acuerdo con formas de realización de la presente invención;

la figura 4 es un diagrama de flujo que representa un método ilustrativo de sincronización de anotaciones entre un documento impreso y un documento electrónico, de acuerdo con formas de realización de la presente invención; y

la figura 5 es un diagrama de flujo que representa un método ilustrativo de detección de anotaciones en una parte de documento impreso de acuerdo con formas de realización de la presente invención.

Aunque la presente invención está abierta a varias modificaciones y formas alternativas, en los dibujos se han mostrado a título de ejemplo formas de realización específicas y las mismas se describen de forma detallada posteriormente. Sin embargo, la presente invención no se limita a las formas de realización particulares descritas. Por el contrario, la presente invención está destinada a abarcar todas las modificaciones, equivalentes, y alternativas que se sitúen dentro del ámbito de la presente invención según queda definido por las reivindicaciones adjuntas.

Aunque el término “bloque” se puede usar en la presente para connotar diferentes elementos de métodos ilustrativos utilizados, el término no debe interpretarse de manera que implique ninguna exigencia de diversas etapas, u orden particular entre ellas, dadas a conocer en la presente, a no ser que se haga referencia explícitamente al orden de etapas individuales y solamente en ese caso.

Descripción detallada

En formas de realización, un lector tal como un estudiante, un doctor o un abogado, puede que desee mantener al mismo tiempo una versión impresa de un documento y una versión electrónica del mismo, o puede que desee comenzar a usar solamente una versión electrónica de un documento después de haber utilizado la versión impresa. El documento impreso puede incluir un número ilimitado de anotaciones, tales como, por ejemplo, partes resaltadas, subrayadas, notas escritas a mano, marcapáginas, y similares, que el lector puede desear tener a su disposición en la versión electrónica del documento. Formas de realización de la presente invención incluyen un sincronizador que facilita la sincronización de anotaciones, tales como las mencionadas, entre el documento impreso y la versión electrónica, de manera que las anotaciones pueden estar disponibles para el lector en la versión electrónica del documento. Por ejemplo, el lector puede crear imágenes de partes anotadas del documento impreso (por ejemplo, usando una cámara o escáner) y proporcionar dichas imágenes al sincronizador, o el lector puede proporcionar el documento impreso a un servicio de escaneado (el cual, en formas de realización, puede estar asociado al sincronizador). El servicio de escaneado puede usar un dispositivo de escaneado para crear imágenes de las partes anotadas del documento impreso y proporcionar las imágenes resultantes al sincronizador. Adicionalmente, en formas de realización, el sincronizador puede facilitar la navegación de una versión electrónica de un documento impreso tal como, por ejemplo, para facilitar la inserción directa, por parte del lector, de anotaciones digitales en la versión electrónica. Por ejemplo, un lector puede proporcionar una imagen de una parte del documento impreso al sincronizador, el cual usa la imagen para localizar y visualizar una parte correspondiente del documento electrónico, en la cual el lector puede insertar directamente una anotación digital.

La figura 1 representa un entorno de funcionamiento ilustrativo (y, en algunas formas de realización, aspectos de la presente invención) de acuerdo con formas de realización de la presente invención según se ilustra a título de ejemplo. Tal como se muestra en la figura 1, formas de realización del entorno de funcionamiento 100 incluyen un servidor 102 que proporciona una copia 104 (etiquetada como “copia de documento electrónico”) de un documento electrónico 106 (etiquetado como “documento electrónico”) a un dispositivo de lectura 108 y que aloja un sincronizador 126, según se describe de forma más detallada posteriormente. En formas de realización, el sincronizador 126 puede ser alojado por el dispositivo de lectura 108 u otro dispositivo informático, y el servidor 102 puede actuar simplemente como un repositorio para el documento electrónico 106.

Como su nombre implica, el dispositivo de lectura 108 es, en formas de realización, lo que usa el lector para leer documentos electrónicos, e incluye un módulo de visualización 110 en el cual se puede visualizar la copia 104 del documento electrónico 106. En formas de realización, el documento electrónico 106 y la copia 104 son versiones electrónicas de un documento impreso (no mostrado) y pueden estar en su forma original, sin anotaciones, o en una forma con anotaciones. Una versión electrónica de un documento impreso puede incluir el mismo contenido que el documento impreso o un contenido sustancialmente similar, aunque también puede incluir contenido diferente. Por ejemplo, una versión electrónica de un documento impreso puede incluir una publicación (por ejemplo, edición) actualizada del documento impreso, una versión con anotaciones del documento impreso, y similares. Los ejemplos de documentos (tanto impresos como electrónicos) incluyen libros, artículos, resoluciones judiciales, recopilaciones de leyes, tratados, notas a pie de página, notas de referencia, notas de traducciones, y similares.

En formas de realización, un usuario (no mostrado) descarga la copia 104 del documento electrónico 106 utilizando el dispositivo de lectura 108, para acceder al servidor 102 a través de una red de comunicaciones 112 tal como, por ejemplo, una red de área local, una red empresarial, Internet, o similares. La copia 104 del documento electrónico 106 también se puede proporcionar al dispositivo de lectura 108 por medio de un dispositivo de memoria extraíble tal como, por ejemplo, un disco compacto, una unidad de almacenamiento flash, o similares. De acuerdo con formas de realización, el documento electrónico 106 (y la copia 104 del mismo) se puede materializar en uno o más archivos usando un número ilimitado de diversos formatos tales como, por ejemplo, DjVu, EPUB®, FictionBook Kindle®, Reader de Microsoft®, eReader®, Plucker, texto plano en ASCII, UNICODE, lenguajes de marcado, un formato de documento independiente de la plataforma, tal como el formato de documento portátil (PDF), y similares. El documento electrónico 106 (y/o la copia 104 del mismo) también se puede materializar en el formato Thomson Reuters ProView®, disponible en Thomson Reuters de Nueva York, Nueva York. Los ejemplos de lenguajes de marcado, y archivos de lenguaje de marcado correspondientes, incluyen Lenguaje de Marcado de Hipertexto (HTML), Lenguaje de Marcado Extensible (XML), Lenguaje de Marcado de Hipertexto Extensible (XHTML), y similares.

Tal como se muestra en la figura 1, el dispositivo de lectura 108 incluye un procesador 114 y una memoria 116. De acuerdo con formas de realización, el dispositivo de lectura 108 es un dispositivo informático y puede adoptar, por ejemplo, la forma de un dispositivo informático especializado o un dispositivo informático de propósito general, tal como un ordenador

personal, una estación de trabajo, un asistente personal digital (PDA), un teléfono móvil, un teléfono inteligente, una tableta, un ordenador portátil tipo *notebook*, o similares. En la memoria 116 hay almacenado un componente de lector de libros electrónicos 118. En formas de realización, el procesador 114 ejecuta el componente de lector de libros electrónicos 118, lo cual puede provocar que por lo menos una parte de la copia 104 del documento electrónico 106 se visualice en el módulo de visualización 110. El componente de lector de libros electrónicos 118 también puede facilitar otras operaciones e interacciones asociadas a la copia 104, tales como, por ejemplo, inserción de anotaciones digitales, búsqueda, marcas en libros, y similares, según se explica de forma más detallada posteriormente. En formas de realización, el componente de lector de libros electrónicos 118 puede acceder al servidor 102 para provocar que por lo menos una parte del documento electrónico 106 se visualice en el módulo de visualización 110.

Tal como se muestra en la figura 1, el servidor 102 incluye un procesador 122 y una memoria 124. El sincronizador 126 puede estar almacenado en la memoria 124. En formas de realización, el procesador 122 ejecuta el sincronizador 126, el cual puede facilitar la navegación de un documento electrónico (o una copia 104 del mismo, por ejemplo, interaccionando con un dispositivo de lectura 108) y la sincronización de anotaciones entre un documento impreso y un documento electrónico 106 (y/o una copia 104 del mismo). El documento electrónico 106 se puede almacenar en unos medios de almacenamiento de contenido 128 en la memoria 124. En formas de realización, para facilitar la sincronización de anotaciones, se proporciona una imagen 107 de una parte de un documento impreso (a la que se hace referencia en la presente como “parte de documento impreso”) al servidor 102, el cual también puede almacenar la imagen 107 en los medios de almacenamiento de contenido 128.

De acuerdo con formas de realización, una parte de documento impreso puede incluir una o más anotaciones en las proximidades de una parte de texto impreso (a la que se hace referencia en la presente como “parte de texto impreso”) y puede estar contenida en una o más páginas de uno o más documentos impresos. Las anotaciones pueden incluir, por ejemplo, partes resaltadas, subrayadas, notas escritas a mano en un margen o entre líneas de texto impreso, marcadores para libros, marcapáginas adhesivos, y similares. Una parte de texto impreso puede incluir, por ejemplo, uno o más capítulos; uno o más párrafos; una o más líneas; una o más palabras; uno o más caracteres; una o más partes de un capítulo, párrafo, línea, palabra, o carácter; o similares. Adicionalmente, una parte de texto impreso puede incluir un pasaje de texto y una o más notas a pie de página, notas al final, figuras, tablas, o similares, correspondientes.

De acuerdo con formas de realización, la imagen 107 de la parte de documento impreso se puede crear, por ejemplo, usando una cámara 120 integrada con el dispositivo de lectura 108 (por ejemplo, cuando el dispositivo de lectura sea un teléfono inteligente u ordenador de tipo 5 tableta) y se puede proporcionar al servidor 102 por medio de la red de comunicaciones 112. Por ejemplo, un lector puede usar la cámara 120 para fotografiar (es decir, crear una imagen 107 de) una parte con anotaciones de un documento impreso y comunicar la imagen 107 al servidor 102, junto con una solicitud de que una versión electrónica 106 del documento impreso se modifique para incluir anotaciones digitales correspondientes a las anotaciones que 10 aparecen en la imagen 107. El servidor 102 puede responder a la solicitud insertando las anotaciones digitales y proporcionando una copia 104 del documento electrónico 106, que tiene las anotaciones digitales correspondientes, al dispositivo de lectura 108, o dando órdenes al componente de lector de libros electrónicos 118 en el dispositivo de lectura 108 para insertar las anotaciones.

15 En formas de realización, la imagen 107 también se puede crear usando un dispositivo de escaneado 130, tal como un escáner industrial, o cualquier otro tipo de dispositivo de formación de imágenes (no mostrado). Por ejemplo, un individuo o una entidad (por ejemplo, una biblioteca, una escuela, un despacho de abogados, o similares) puede proporcionar 20 documentos impresos con anotaciones a un proveedor de servicios asociado al servidor 102. El proveedor de servicios puede utilizar un escáner industrial 130, por ejemplo, para escanear grandes cantidades de documentos, libros completos, o similares, y proporcionar las imágenes 107 creadas a partir del proceso de escaneado directamente al servidor 102. En formas de realización, un servicio de escaneado puede utilizar un dispositivo de escaneado 130 y 25 proporcionar las imágenes resultantes 107 al servidor 102 por medio de la red de comunicaciones 112.

De acuerdo con formas de realización, para facilitar la sincronización de anotaciones, el sincronizador 126 recupera, o alternativamente accede a, por lo menos una parte de la imagen 30 107 de los medios de almacenamiento de contenido 128, y recupera, o alternativamente accede a, por lo menos una parte de una versión electrónica (por ejemplo, documento electrónico 106) del documento impreso. El sincronizador 126 identifica una parte de texto electrónico (a la que se hace referencia en la presente como “parte de texto electrónico”) del documento electrónico 106, que se corresponde con la parte de texto impreso capturada en la 35 imagen 107. De acuerdo con formas de realización, una parte de texto electrónico se corresponde con una parte de texto impreso si las dos partes de texto son textualmente

similares. La expresión similitud textual puede referirse, por ejemplo, a un grado de similitud entre dos partes de texto y se puede definir, por ejemplo, en el contexto de medidas estadísticas, relaciones, o similares. Por ejemplo, dos partes de texto pueden ser textualmente similares si las mismas presentan un cierto número (por ejemplo, por comparación con otras partes de texto adyacentes) de caracteres coincidentes, n-gramas de caracteres, o similares.

Adicionalmente, en formas de realización, el sincronizador 126 se puede configurar para analizar la imagen 107 con el fin de detectar anotaciones en la parte de documento impreso y también se puede configurar para interpretar anotaciones detectadas. Por ejemplo, el sincronizador 126 puede detectar una anotación dentro de una imagen 107 de una parte de documento impreso y puede determinar un tipo de anotación (por ejemplo, resaltado, subrayado, y similares) de la anotación detectada. De acuerdo con formas de realización, se pueden utilizar uno o más revisores (por ejemplo, por medio de un modelo de *crowd-sourcing*) para facilitar la detección y/o interpretación de anotaciones, así como la creación, modificación, y/o verificación de anotaciones digitales. Por ejemplo, el servidor 102 puede proporcionar una imagen de una anotación escrita a mano, a un dispositivo de revisión 132 (por ejemplo, por medio de la red de comunicaciones 112), de manera que un revisor pueda ayudar a la creación de la anotación digital detectando, interpretando y/o transcribiendo la anotación escrita a mano, en texto digital, en el cual se pueden realizar búsquedas utilizando el dispositivo de lectura 108. En formas de realización, revisores adicionales pueden verificar las interpretaciones, del primer revisor, correspondientes a la anotación escrita a mano. Se pueden usar plataformas de *crowd-sourcing* para interactuar con revisores y las mismas pueden incluir plataformas de *crowd-sourcing* integradas con el servidor 102 o plataformas independientes tales como, por ejemplo, Amazon Mechanical Turk®, proporcionado por Amazon.com® Inc. de Seattle, Washington, Estados Unidos. También se pueden utilizar algoritmos de inteligencia artificial para interpretar, modificar y/o verificar anotaciones digitales.

El sincronizador 126 puede insertar una anotación digital correspondiente a la anotación detectada, en el documento electrónico 106, y, en formas de realización, el componente de lector de libros electrónicos 118 puede insertar la anotación digital en la copia 104 del documento electrónico 106. Por ejemplo, el sincronizador 126 puede proporcionar una instrucción al componente de lector de libros electrónicos 118, que provoque que el componente de lector de libros electrónicos 118 inserte la anotación digital en la copia 104 del documento electrónico 106. Los ejemplos de anotación digital incluyen, aunque sin carácter limitativo, una imagen de una anotación detectada, texto electrónico, un objeto formateador (por ejemplo, código que provoque que una parte de texto electrónico reproducido incluya una

parte resaltada, una parte subrayada, o similares), un marcador de libro electrónico, un marcapáginas digital, un hipertexto, un conjunto de instrucciones para modificar una parte del documento electrónico 106 (o copia 104 del mismo), y similares.

5 Formas de realización de la presente invención pueden facilitar también diversas operaciones para gestionar anotaciones digitales. Por ejemplo, el sincronizador 126, y/o el componente de lector de libros electrónicos 118, se pueden configurar para eliminar una anotación digital de un documento electrónico 106 (o copia 104 del mismo) que había sido insertada previamente durante un proceso de sincronización, y que ya no se encuentra en el documento impreso
10 correspondiente. Adicionalmente, el sincronizador 126, y/o componente de lector de libros electrónicos 118, se puede configurar para distinguir entre anotaciones digitales que se insertaron durante un proceso de sincronización (a las que se hace referencia en la presente como anotaciones digitales migradas) y anotaciones digitales que se añadieron directamente en el documento electrónico 106 (o copia 104 del mismo) (a las que se hace referencia en la
15 presente como anotaciones digitales directas). En formas de realización, una anotación digital puede incluir un atributo que indica específicamente si la anotación digital se insertó durante un proceso de sincronización o fue añadida directamente por un lector. El atributo se puede representar mediante un valor de una etiqueta que está asociada al documento electrónico 106 y el sincronizador puede determinar si la anotación es una anotación digital migrada o una
20 anotación digital directa determinando el valor de la etiqueta. De esta manera, formas de realización de la presente invención pueden facilitar la eliminación de anotaciones digitales migradas, al mismo tiempo que reduciendo al mínimo la eliminación no intencionada de anotaciones digitales directas.

25 Los valores de etiquetas también se pueden asociar a migraciones de anotaciones particulares. Por ejemplo, un lector puede colocar un primer pólito relacionado con un primer contenido (por ejemplo, un primer caso jurídico) en un documento impreso. Cuando en una versión electrónica del documento impreso se inserta una primera anotación digital, correspondiente al primer pólito, un valor de etiqueta asociado a la anotación digital puede proporcionar información sobre
30 la instancia de la migración. El valor de etiqueta puede indicar, por ejemplo, que la primera anotación digital está relacionada con el primer caso (por ejemplo, cuando el lector pueda especificar el valor de etiqueta), que se creó en un momento particular, o similares. Posteriormente, el lector podría eliminar el primer pólito del documento impreso y sustituirlo por un segundo pólito relacionado con un segundo contenido (por ejemplo, un segundo caso
35 jurídico). En formas de realización, en el documento electrónico también se puede insertar una segunda anotación digital, correspondiente al segundo pólito, y la misma puede tener un valor

de etiqueta asociado que indique, por ejemplo, que está relacionada con el segundo caso jurídico. Al lector se le puede presentar una opción de mantener o eliminar la primera anotación digital, personalizar valores de etiqueta asociados a la primera y/o la segunda anotaciones digitales, o similares.

5 Formas de realización de la invención facilitan también la manipulación de anotaciones digitales. Por ejemplo, las anotaciones digitales se pueden almacenar como elementos independientes adjuntos al documento electrónico, que se pueden manipular, o las anotaciones digitales se pueden integrar dentro del documento electrónico, el cual se puede manipular.
10 Adicionalmente, el sincronizador 126, y/o el componente de lector de libros electrónicos 118, se puede configurar para ajustar las posiciones de anotaciones digitales, de tal manera que anotaciones digitales nuevas no oculten anotaciones digitales existentes. En formas de realización, el sincronizador 126, y/o el componente de lector de libros electrónicos 118 pueden permitir que un lector busque, manipule, cambie de posición, edite, borre, o gestione de otra
15 manera anotaciones digitales.

Formas de realización de la presente invención también pueden facilitar la visualización selectiva de anotaciones digitales. Por ejemplo, si la copia 104 del documento electrónico con anotaciones 106 se usa durante una sesión de un tribunal, podría resultar deseable presentar
20 una sección de la copia 104 al abonado contrario, al juez, y/o al jurado sin visualizar una o más anotaciones digitales contenidas en ella. Así, el componente de lector de libros electrónicos 118 puede incluir una opción para ocultar una o más anotaciones digitales cuando se visualice la copia 104. Para facilitar esto, una anotación digital puede incluir una etiqueta que permita mostrar u ocultar la anotación digital, sobre la base del valor de la etiqueta. Por ejemplo, un
25 lector puede especificar que anotaciones digitales que presenten un cierto valor de etiqueta (por ejemplo, anotaciones digitales relacionadas con un primer caso jurídico, según se ha descrito en el ejemplo anterior) sean ocultadas. El valor de la etiqueta también puede posibilitar el formateo condicional de una anotación digital. Por ejemplo, el tamaño, la forma, el formato de archivo, y/o la disposición de una anotación digital se pueden ajustar sobre la base de un
30 valor de la etiqueta, el cual puede ser estático o dinámico y se puede asignar basándose en características de anotaciones digitales cercanas, limitaciones de memoria, área de pantalla disponible, capacidades del dispositivo de lectura, o similares. Adicionalmente, se pueden generar valores de etiqueta de manera manual o automática. Por ejemplo, el componente de lector de libros electrónicos 118 puede asignar un valor particular a una etiqueta, basándose en
35 un acontecimiento o condición, y el componente de lector de libros electrónicos 118 puede hacer que una opción seleccionable (por ejemplo, un botón o icono) sea presentado en el

módulo de visualización 110 de manera que, al producirse la recepción de una selección de la opción, una o más anotaciones, que de otro modo pueden visualizarse, entonces no lo sean.

De acuerdo con formas de realización, varios componentes del entorno de funcionamiento 100, ilustrado en la figura 1, se pueden implementar en uno o más dispositivos informáticos. Por ejemplo, cada uno del servidor 102, el dispositivo de lectura 108, el dispositivo de escaneado 130, y el dispositivo de revisión 132 puede ser, o incluir, uno o más dispositivos informáticos. Un dispositivo informático puede incluir cualquier tipo de dispositivo informático adecuado para implementar formas de realización de la invención. Los ejemplos de dispositivos informáticos incluyen “estaciones de trabajo”, “servidores”, “ordenadores portátiles”, “ordenadores de sobremesa”, “ordenadores de tipo tableta”, “dispositivos de mano”, “lectores de libros electrónicos”, y similares, contemplándose todos ellos dentro del alcance de la figura 1 y haciendo referencia a varios componentes del entorno de funcionamiento 100.

En formas de realización, un dispositivo informático incluye un bus, de manera directa y/o indirecta, acopla los siguientes dispositivos: un procesador, una memoria, un puerto de entrada/salida (I/O), un componente de I/O, y una fuente de alimentación. En el dispositivo informático se puede incluir también un número ilimitado de componentes adicionales, componentes diferentes, y/o combinaciones de componentes. El bus representa lo que puede ser uno o más buses (tal como, por ejemplo, un bus de direcciones, un bus de datos, o una combinación de los mismos). De manera similar, en formas de realización, el dispositivo informático puede incluir varios procesadores, varios componentes de memoria, varios puertos de I/O, varios componentes de I/O, y/o varias fuentes de alimentación. Adicionalmente, un número ilimitado de estos componentes o sus combinaciones se pueden distribuir y/o duplicar sobre un número de dispositivos informáticos.

En formas de realización, las memorias 116 y 124 incluyen soportes legibles por ordenador, en forma de memoria volátil y/o no volátil y pueden ser extraíbles, no extraíbles, o una combinación de los mismos. Los ejemplos de soportes incluyen Memoria de Acceso Aleatorio (RAM); Memoria de Solo Lectura (ROM); Memoria de Solo Lectura, Programable y Borrable Electrónicamente (EEPROM); memoria flash; soportes ópticos u holográficos; casetes magnéticos, cinta magnética, dispositivos de almacenamiento de disco magnético u otros dispositivos de almacenamiento magnético; transmisiones de datos; o cualquier otro medios que se pueda usar para codificar información y al que se pueda acceder por medio de un dispositivo informático tal como, por ejemplo, memoria de estado cuántico, y similares. En formas de realización, las memorias 116 y 124 almacenan instrucciones ejecutables por

ordenador para provocar que los procesadores 114 y 122, respectivamente, ejecuten aspectos de formas de realización de métodos y procedimientos que se describen en la presente. Las instrucciones ejecutables por ordenador pueden incluir, por ejemplo, código de ordenador, instrucciones utilizables en máquinas, y similares, tales como, por ejemplo, componentes de programa con capacidad de ser ejecutados por uno o más procesadores asociados a un dispositivo informático. Los ejemplos de dichos componentes de programa incluyen el componente de lector de libros electrónicos 118 y el sincronizador 126. Parte o la totalidad de la funcionalidad que se contempla en la presente también se puede implementar en hardware y/o microprogramas.

10 El entorno de funcionamiento 100 ilustrativo que se muestra en la figura 1 no está destinado a sugerir ninguna limitación en cuanto al alcance de uso o funcionalidad de formas de realización de la presente invención. Tampoco debe interpretarse el entorno de funcionamiento 100 ilustrativo de manera que presente dependencia o exigencia alguna en relación con cualquier componente individual o combinación de componentes ilustrados en la presente.

La figura 2 es un diagrama esquemático que representa un funcionamiento ilustrativo del sincronizador 126 de acuerdo con formas de realización de la presente invención. Tal como se muestra en la figura 2, el sincronizador 126 puede incluir un componente de detección 202, un componente de comparación 204, y un componente de anotación digital 206. De acuerdo con formas de realización, uno cualquiera o más de los componentes 202, 204, y 206 del sincronizador 126 representado en la figura 2 pueden compartir recursos, o estar integrados, con diversos de los otros componentes representados en el mismo lugar (y/o componentes no ilustrados). Adicionalmente, en formas de realización, operaciones de los componentes 202 y 204 se pueden llevar a cabo en cualquier orden, ciclo, combinación, o similar. Por ejemplo, el componente de comparación 204 se puede utilizar para identificar una parte de documento electrónico que se corresponde con una parte de documento impreso antes de la utilización del componente de detección 202 para detectar una anotación en la parte de documento impreso. Adicionalmente, el componente de detección 202 se puede utilizar para detectar una anotación en una parte de documento impreso antes de la utilización del componente de comparación 204 para identificar la parte de documento electrónico correspondiente en la cual se va a insertar una anotación digital correspondiente. Adicionalmente, uno cualquiera o más de los componentes 202, 204, y 206 pueden residir en el servidor 102 o el dispositivo de lectura 108, o pueden estar distribuidos entre el servidor 102 y el dispositivo de lectura 108.

35 De acuerdo con formas de realización, el componente de detección 202 detecta una anotación

218 en una parte de documento impreso 210. Por ejemplo, el componente de detección 202 puede recibir una imagen 107 de la parte de documento impreso 210 (por ejemplo, de la memoria 124) y puede ejecutar uno o más procedimientos para detectar la anotación 218. En formas de realización, el componente de detección 202 también puede identificar un tipo de anotación correspondiente a la anotación 218. Los ejemplos de tipos de anotación incluyen texto, resaltados, subrayados, marcapáginas, marcadores para libros, y similares. Se puede usar un número ilimitado de tipos diferentes de procedimientos para detectar anotaciones y/o identificar tipos de anotación. Los ejemplos de dichos procedimientos pueden incluir procedimientos de reconocimiento de escritura manual, procedimientos de reconocimiento óptico de caracteres (OCR), procedimientos de comparación de mapas de bits, modelos de idioma estadísticos, clasificadores estadísticos, redes neuronales, *crowd-sourcing*, y similares.

Por ejemplo, el componente de detección 202 puede analizar la imagen 107 (por ejemplo, examinando píxeles o bloques de píxeles) para establecer patrones asociados a la parte de texto impreso 216 y para detectar una característica, o características, anómalas que pueden representar la anotación 218. Los ejemplos de dichas características anómalas pueden incluir instancias de colores diferentes (por ejemplo, asociados a un resaltado), instancias de formas y bordes irregulares (por ejemplo, asociados a notas escritas a mano o subrayados, instancias de texto oculto o formas geométricas que presentan diferentes tonos de color (por ejemplo, asociadas a pósts, marcapáginas, o marcadores de libros), y similares. Por ejemplo, una parte de texto impreso se puede haber resaltado con un marcador de texto amarillo, el cual se puede detectar como una característica de color amarillo que se solapa parcialmente con una parte de texto impreso. En formas de realización, el componente de detección 202 puede usar técnicas estadísticas para determinar una probabilidad de que una anomalía detectada represente una anotación 218.

De acuerdo con formas de realización, el componente de detección 202 puede detectar la anotación 218 en la parte de documento impreso 210 comparando la imagen 107 de la porción de documento impreso 210 con por lo menos una porción del documento electrónico 106. Por ejemplo, cuando el componente de comparación 204 se utiliza antes del componente de detección 202, el componente de comparación 204 puede proporcionar una indicación, al componente de detección 202, sobre una parte de texto electrónico 214 que se corresponde con la parte de texto impreso 216. El componente de detección 202 puede, por ejemplo, acceder, o crear, un mapa de bits de una parte correspondiente del documento electrónico 106 y comparar ese mapa de bits con un mapa de bits (por ejemplo, la imagen 107) de la parte de documento impreso 210 para identificar diferencias entre los dos mapas de bits. Una diferencia

en los mapas de bits puede representar, por ejemplo, la presencia de la anotación 218 que no está presente en el documento electrónico 106. Una diferencia en los mapas de bits también puede representar la presencia de una anotación digital (no mostrada) en el documento electrónico 106 que no está presente en la parte de documento impreso 210.

5

En formas de realización, el componente de detección 202 puede utilizar un procedimiento de zonificación de imágenes para fraccionar la imagen 107 en zonas. Un procedimiento de zonificación de imágenes puede definir una o más zonas de texto correspondientes a texto impreso, y una o más zonas candidatas correspondientes a áreas que pueden incluir anotaciones tales, por ejemplo, partes de un margen. El procedimiento de zonificación de imágenes puede definir un margen como una región que se encuentra a la izquierda o derecha, respectivamente, de una zona de texto situada más a la izquierda, o más a la derecha. En formas de realización, una zona de texto también puede ser una zona candidata. Se puede aplicar un proceso de reconocimiento de escritura manual a una zona en un intento de reconocer un escrito a mano. Si el proceso de reconocimiento de escritura manual tiene éxito en el reconocimiento del escrito a mano dentro de la zona, el componente de detección 202 puede identificar la escritura manual reconocida como una anotación. El reconocimiento de la escritura manual puede incluir la conversión de la escritura manual en texto electrónico en el cual se pueden realizar búsquedas, y el cual se puede insertar en el documento electrónico como una anotación digital directa.

10
15
20

En formas de realización, puede que una parte de escritura manual no sea reconocible por el procedimiento de reconocimiento de escritura manual o puede que la zona no incluya un escrito a mano, y, como consecuencia, el procedimiento de reconocimiento de escritura manual puede no tener éxito con el reconocimiento de la escritura manual. Adicionalmente, puede que no haya disponible un procedimiento de reconocimiento de escritura manual o que el mismo resulte poco práctico (por ejemplo, debido a limitaciones de memoria). En tales casos, se pueden llevar a cabo otros tipos de análisis sobre la zona para determinar si la misma contiene escritura manual no reconocible u otros tipos de anotaciones. Por ejemplo, se puede aplicar un procedimiento OCR para intentar identificar anotaciones dentro de la zona (por ejemplo, etiquetando características detectadas dentro de la zona de las cuales no se puede establecer una correspondencia con un carácter textual por medio del procedimiento OCR), según se describe de forma más detallada posteriormente en referencia a la figura 5. En formas de realización, se puede usar un procedimiento de detección de regiones para identificar regiones de color homogéneo superpuestas sobre, o adyacentes a, texto impreso detectado, lo cual puede facilitar la identificación de partes resaltadas, subrayadas, y similares. Se puede utilizar

25
30
35

un número ilimitado de clasificadores estadísticos, heurística, redes neuronales, y similares, para detectar anotaciones, y herramientas tales como las mencionadas se pueden mejorar de forma periódica o continua utilizando técnicas de aprendizaje por máquina. Tal como se ha indicado anteriormente, el componente de detección 202 también puede determinar el tipo de anotación 218 que detecta. Por ejemplo, el componente de detección 202 puede determinar que una anotación detectada 218 en la parte de documento impreso 210 es de un tipo de anotación particular basándose, por lo menos parcialmente, en tablas de consulta, características de anotación 218, retroalimentación proveniente de un revisor, salidas de clasificadores, y similares.

10 Tal como se ha indicado anteriormente, el componente de comparación 204 está configurado para acceder a por lo menos una parte de una imagen 107 de una parte de documento impreso 210 de una memoria (por ejemplo, la memoria 124 mostrada en la figura 1) y para acceder a por lo menos una parte de un documento electrónico 106 de una memoria. El componente de comparación 204 identifica una parte de texto electrónico 214 en el documento electrónico 106, que se corresponde con (por ejemplo, que es textualmente similar a) una parte de texto impreso 216 de la parte de documento impreso 210. El componente de comparación 204 puede utilizar algoritmos que incorporan técnicas de OCR, técnicas de análisis sintáctico, y/o similares, para identificar una parte de texto electrónico 214 que es textualmente similar a una parte de texto impreso 216.

En formas de realización, una versión electrónica 106 de un documento impreso puede no incluir texto idéntico al del documento impreso. Por ejemplo, una parte de texto impreso 216 puede incluir un pasaje de texto y tres notas correspondientes a pie de página, mientras que la parte de texto electrónico 214 correspondiente puede incluir el pasaje de texto y cuatro notas correspondientes a pie de página (por ejemplo, la cuarta nota a pie de página puede exponer un caso decidido recientemente) donde, por ejemplo, la versión electrónica es una edición más reciente del documento. Para facilitar la identificación de partes de texto correspondientes de una manera que tenga en cuenta variaciones textuales tales como las mencionadas, se pueden configurar algoritmos utilizados por el componente de comparación 204 para evaluar similitudes textuales entre las partes de texto 214 y 216. Tal como se ha mencionado previamente, una similitud textual se puede referir a un grado de solapamiento entre caracteres reconocidos (por ejemplo, por medio de OCR) en la parte de texto impreso 216 y caracteres en la parte de texto electrónico 214. Por ejemplo, el componente de comparación 204 puede usar técnicas de búsqueda de n-gramas para identificar un conjunto de líneas de un documento electrónico 106 que incluye el número mayor de n-gramas que coinciden con n-gramas del texto reconocido de

la parte de texto impreso 216. En formas de realización, el componente de comparación 204 puede utilizar un número ilimitado de técnicas de comparación estadística para evaluar la similitud textual.

5 En algunos casos, la parte de texto impreso 216 puede incluir únicamente partes de palabras y/o caracteres. Por ejemplo, la imagen 107 se puede capturar de tal manera que la parte de documento impreso 210 completa no quede capturada dentro de la imagen 107, o un pólito u otro objeto puede ocultar parte de la parte de texto impreso 216 cuando se capture la imagen 107. En formas de realización, la comparación idéntica de texto se puede complicar debido a
10 un reconocimiento impreciso de caracteres durante procedimientos de OCR. De acuerdo con formas de realización, el componente de comparación 204 puede utilizar técnicas de poda de caracteres para facilitar la evaluación de similitudes textuales entre las partes de texto 214 y 216. Volviendo brevemente a la figura 3, se ilustra una parte 302 de un documento impreso que tiene una parte de texto impreso 304. Cuando un usuario captura una región 306 de la parte de documento impreso 302 usando un dispositivo de formación de imágenes, la región capturada
15 306 puede incluir solamente una fracción de la parte de texto impreso 304, tal como se ilustra. Cuando la región capturada 306 se digitaliza 308 para crear una imagen 310, la imagen 310 puede incluir líneas de texto, palabras, o caracteres incompletos. Por ejemplo, tal como se muestra en la figura 3, la imagen ilustrada 310 incluye palabras parciales 312, algunas de las
20 cuales incluyen caracteres parciales. Cuando la imagen 310 se procesa 314 usando un procedimiento OCR, el texto electrónico reconocido resultante 316 puede incluir errores de reconocimiento 318 (ilustrados como caracteres subrayados). Las secuencias de caracteres (por ejemplo, palabras parciales) que incluyen los errores de reconocimiento 318 es posible que no se puedan emparejar de manera precisa con entradas de un diccionario de OCR, y
25 pueden reducir potencialmente la eficiencia y/o eficacia del procedimiento de comparación. En formas de realización, se pueden utilizar matrices de sufijos, comparación de expresiones regulares (por ejemplo, utilizando caracteres de comodín) y/o métodos de comparación de cadenas aproximadas (por ejemplo, algoritmos de Distancia de Edición) para interpretar las secuencias de caracteres que contienen los errores de reconocimiento 318.

30 Adicionalmente, formas de realización de la invención incluyen eliminar 320 las secuencias de caracteres que contienen los errores de reconocimiento 318 para crear texto electrónico reconocido y podado 322. El texto electrónico reconocido y podado 322, el cual incluye el texto electrónico reconocido 316 sin los errores de reconocimiento 318, se puede usar para facilitar
35 un procedimiento de comparación. El procedimiento de comparación puede implicar la evaluación de una similitud textual (en lugar de una identidad) y, por lo tanto, una parte de texto

reconocido que ha sido podada puede seguir siendo útil en la identificación de una parte de texto electrónico correspondiente. Por ejemplo, se pueden usar una o más consultas de búsqueda para buscar en el documento electrónico porciones de texto electrónico correspondientes, y las mismas pueden dar como resultado la recuperación de partes de texto electrónico que estén clasificadas de manera que permitan que una o más de ellas se identifique como partes de texto electrónico correspondientes. La identificación de partes de texto electrónico correspondientes se puede facilitar adicionalmente, con el uso de modelos de *crowd-sourcing*, presentando a un lector o revisor una o más partes de texto electrónico recuperadas (por ejemplo, las partes de texto electrónico recuperadas con la mejor clasificación, partes de texto electrónico con clasificación similar, o similares), y solicitando una entrada proveniente del lector o revisor, que se puede utilizar para identificar o confirmar la parte de texto electrónico correspondiente. El *crowd-sourcing* también se puede usar para la corrección de errores de reconocimiento de OCR. Por ejemplo, las secuencias de caracteres que contienen los errores de reconocimiento 318 (y, en formas de realización, que rodean secuencias de caracteres) se pueden proporcionar a revisores, los cuales pueden verificar la interpretación de OCR o proporcionar sugerencias para una interpretación correcta de las cadenas de caracteres.

De acuerdo con formas de realización de la invención, el componente de comparación 204 puede utilizar varios procedimientos, clasificadores, y similares, para determinar qué fracciones podar de una parte de texto reconocida. Por ejemplo, los errores de reconocimiento pueden aparecer cerca de los límites de una parte de texto reconocido, y las palabras reconocidas cerca de uno o más de los límites que no se encuentran en el diccionario de OCR se pueden podar. Adicionalmente, la imagen de la parte de documento impreso se puede analizar para determinar qué regiones de la imagen incluyen texto impreso, lo cual también puede resultar útil para facilitar procedimientos de poda.

Volviendo a la figura 2, el sincronizador 126 puede incluir un componente de anotación digital 206 que facilite la inserción de una anotación digital 220 en el documento electrónico 106 (o en una copia 104 del mismo). La inserción de una anotación digital 220 en un documento electrónico 106 ó copia 104 puede incluir, por ejemplo, la inserción de texto electrónico (por ejemplo, como una anotación digital directa proporcionada por un lector, como una anotación digital migrada, o similares), la adición de código correspondiente a la anotación digital 220 a un lenguaje de marcado extensible, la asociación de un archivo de anotaciones a un archivo de documentos electrónicos, la incorporación de la anotación digital 220 dentro de un archivo de documentos electrónicos, o similares. El componente de anotación digital 206 puede facilitar la

inserción de la anotación digital 220 insertando la anotación digital 220 en el documento electrónico 106 (y/o una copia 104 del mismo), el cual a continuación se puede proporcionar al dispositivo de lectura 108.

5 De acuerdo con formas de realización, el componente de anotación digital 206 inserta la anotación digital 220 accediendo al texto electrónico del documento electrónico con el uso de una interfaz de programación de aplicaciones (API). Por ejemplo, el componente de anotación digital 206 puede insertar una anotación digital en un documento PDF usando la API de Desarrollador (*Developer*) de Adobe®, disponible en Adobe Systems, Inc., de San José,
10 California. La anotación digital 220 se puede insertar en un documento electrónico basado en HTML proporcionando una superposición de HTML o proporcionando un esquema o archivo de metadatos e insertando, en el archivo de HTML, un puntero al esquema o archivo de metadatos. La anotación digital 220 se puede generar usando un procedimiento de reconocimiento de escritura manual y la misma se puede insertar en el documento electrónico
15 como texto electrónico en el cual se pueden realizar búsquedas. Adicionalmente, el componente de anotación digital 206 puede facilitar la inserción de la anotación digital 220 proporcionando la anotación digital 220 al componente de lector de libros electrónicos 118, el cual inserta la anotación digital 220 en la copia 104 del documento electrónico 106. El componente de anotación digital 206 puede proporcionar una instrucción que se comunica al
20 componente de lector de libros electrónicos 118 y que provoca que el componente de lector de libros electrónicos 118 inserte la anotación digital 220 en la copia 104 del documento electrónico 106.

Tal como se ha descrito anteriormente, un sincronizador puede facilitar la navegación de un
25 documento electrónico y/o la sincronización de anotaciones entre un documento impreso y un documento electrónico. La figura 4 es un diagrama de flujo que representa un método ilustrativo 400 de sincronización de anotaciones entre un documento impreso y un documento electrónico, de acuerdo con formas de realización de la presente invención. Formas de realización del método 400 incluyen la recepción de una imagen de una parte de documento
30 impreso (bloque 410). La parte de documento impreso puede incluir una anotación en las proximidades de una parte de texto impreso. Un sincronizador (por ejemplo, el sincronizador 126 mostrado en la figura 1) puede recibir la imagen de un dispositivo de lectura (por ejemplo, el dispositivo de lectura 108 mostrado en la figura 1), un dispositivo de escaneado (por ejemplo, el dispositivo de escaneado 130 mostrado en la figura 1), o similares, y puede almacenar la
35 imagen en una memoria (por ejemplo, la memoria 124 mostrado en la figura 1).

El sincronizador puede recuperar la imagen y un documento electrónico de la memoria (bloque 420). El sincronizador puede usar técnicas de modelado de lenguaje, OCR, o similares para identificar el documento electrónico que se corresponde con la parte de texto impreso. Esto puede incluir el uso de un clasificador de idiomas para determinar el idioma en el que está escrito el libro. En formas de realización del método 400, el sincronizador identifica una parte de texto electrónico dentro del documento electrónico, que se corresponde con la parte de texto impreso (bloque 430). Por ejemplo, el sincronizador puede incluir un componente de comparación (por ejemplo, componente de comparación 202 mostrado en la figura 2) que identifica una parte de texto electrónico que es textualmente similar a la parte de texto impreso. El componente de comparación puede llevar a cabo una búsqueda dentro del documento electrónico, en la cual líneas de la parte de texto impreso son consultas de expresiones que se aplican con respecto a una forma indexada del documento electrónico. El documento electrónico indexado puede incluir, por ejemplo, matrices de sufijos. Múltiples consultas de expresiones que conducen a resultados de recopilación con la mejor clasificación se pueden usar para identificar la parte de texto electrónico correspondiente.

Formas de realización del método 400 incluyen también la detección de una anotación en la parte de documento impreso (bloque 440). Por ejemplo, un componente de detección (por ejemplo, el componente de detección 204 mostrado en la figura 2) puede detectar la anotación analizando una imagen de la parte de documento impreso. Se pueden detectar anotaciones usando procedimientos de OCR, procedimientos de reconocimiento de escritura manual, clasificadores, o una combinación de los mismos, tal como se describe de forma más detallada posteriormente, en referencia a la figura 5. Tal como se muestra en la figura 4, en el documento electrónico y/o en una copia del documento electrónico (bloque 450) se inserta una anotación digital, correspondiente a la anotación detectada. Tal como se ha descrito anteriormente, la anotación digital se puede insertar en el documento electrónico por medio del sincronizador, o se puede insertar en una copia del documento electrónico por medio de un componente de lector de libros electrónicos (por ejemplo, el componente de lector de libros electrónicos 118 mostrado en la figura 1). La anotación digital se puede insertar en las proximidades de la parte de texto electrónico correspondiente, en una posición que se corresponda por lo menos sustancialmente con la posición de la anotación detectada en la parte de documento impreso.

La figura 5 es un diagrama de flujo que representa un método ilustrativo 500 de detección de una anotación en una parte de documento impreso analizando una imagen de la parte de documento impreso, de acuerdo con formas de realización de la presente invención. Formas de realización del método ilustrativo 500 incluyen la definición de una zona candidata (bloque

510), por ejemplo, fraccionando la imagen en varias zonas, las cuales pueden incluir zonas de texto, zonas candidatas, y similares. En algunos casos, las zonas de texto también pueden ser zonas candidatas (por ejemplo, en los casos en los que las anotaciones pueden estar presentes en o entre líneas de texto impreso). Un componente de detección (por ejemplo, el componente de detección 204 mostrado en la figura 2) puede fraccionar la imagen en zonas basándose en un patrón geométrico (por ejemplo, rectangular), en regiones de la imagen que contengan características homogéneas, o similares. Se pueden aplicar uno o más clasificadores estadísticos para distinguir entre regiones de la imagen que es más probable que incluyan texto impreso y regiones que es más probable que contenga anotaciones. Formas de realización de la invención incluyen entonces la ejecución de una o más de las siguientes etapas (520 a 580) para cada zona definida.

Formas de realización del método 500 incluyen la ejecución de un procedimiento de reconocimiento de escritura manual (bloque 520) sobre una zona candidata, que se puede utilizar, por ejemplo, para detectar anotaciones escritas a mano en la zona candidata. Adicionalmente, el texto escrito a mano en zonas de texto (por ejemplo, entre líneas de texto impreso) se puede identificar usando clasificadores que distinguen entre escritura a mano y texto impreso. Los ejemplos de dichos clasificadores incluyen clasificadores de máquinas de soporte vectorial (SVM), clasificadores de k vecinos más cercanos (K-NN), discriminadores de Fischer, redes neuronales, clasificadores de distancia mínima, y similares.

Sobre la zona candidata se puede ejecutar un procedimiento OCR (530). El procedimiento OCR se puede usar para extraer características de la zona candidata, y para determinar si las características extraídas están asociadas a texto impreso. Si el procedimiento OCR no da como resultado la detección de ningún texto impreso, se puede deducir que la zona candidata puede que contenga anotaciones. De manera similar, si el procedimiento OCR detecta texto impreso en solamente una fracción de la zona candidata, se puede deducir que otras fracciones de la zona candidata puede que contengan anotaciones. El procedimiento OCR puede incluir, por ejemplo, extracción de características, comparación matricial, o una combinación de las mismas.

Formas de realización del método 500 incluyen la aplicación de un modelo de lenguaje estadístico a nivel de caracteres (bloque 540) sobre características extraídas por el procedimiento OCR. El modelo de lenguaje estadístico a nivel de caracteres se puede usar para calcular una probabilidad de que una característica extraída incluya una cadena de caracteres que sea típica de un lenguaje particular, por ejemplo, $P(s_c)$, donde s_c es una

secuencia de caracteres. Adicionalmente, se puede aplicar un modelo de lenguaje estadístico a nivel de palabras (bloque 550) sobre características extraídas y el mismo se puede usar, por ejemplo, para calcular una probabilidad de que una característica extraída incluya una secuencia de palabras típica de un lenguaje particular, por ejemplo, $P(s_w)$, donde s_w es una secuencia de palabras. Los modelos de lenguaje se pueden aplicar para facilitar la determinación de si las características extraídas están asociadas probablemente a texto impreso.

Formas de realización del método 500 incluyen además la determinación de información de color asociada a la zona candidata (bloque 560). Por ejemplo, la imagen de la parte del documento impreso puede ser una fotografía de color y el componente de detección puede analizar la fotografía para identificar cada píxel que incluye un color (por ejemplo, diferente de negro o blanco), el color de cada píxel, características de colores de los píxeles (por ejemplo, tono, saturación, intensidad), y similares. De esta manera, por ejemplo, regiones de color en una zona candidata se pueden detectar y usar para facilitar la detección de una anotación y/o identificación de un tipo de anotación correspondiente a una anotación detectada. Por ejemplo, una región relativamente cuadrada de píxeles amarillos en una zona candidata definida en un margen se pueden identificar como un pósito amarillo, mientras que una región rectangular más estrecha de píxeles rosas en una zona de texto se puede identificar como un resaltado. Se pueden usar métodos supervisados de aprendizaje de máquinas para distinguir entre tipos diferentes de anotaciones (por ejemplo, pósitos con respecto a notas marginales escritas a mano con respecto a un resaltado verde con un marcador de texto).

Formas de realización del método 500 incluyen además la comparación de la imagen con una parte de documento electrónico correspondiente (bloque 570). Un mapa de bits de la imagen se puede comparar con un mapa de bits de la parte de documento electrónico correspondiente para evaluar diferencias entre los píxeles de los dos mapas de bits. Por ejemplo, un componente de detección puede superponer el mapa de bits de la imagen sobre el mapa de bits de la parte de documento electrónico y reorientar los mapas de bits superpuestos hasta que se obtenga un número mínimo de diferencias de píxeles en superposición. Las diferencias de píxeles que queden (por ejemplo, píxeles que aparecen en el mapa de bits de la imagen y que no aparecen en el mapa de bits del documento electrónico o viceversa) pueden representar anotaciones. Adicionalmente, se pueden utilizar umbrales de ruido para ignorar diferencias pequeñas (por ejemplo, 1 ó 2 píxeles desconectados) que pueden representar, por ejemplo, partículas de polvo, defectos en la lente de la cámara, o similares. Igual que con cualquiera de los otros procedimientos descritos en la presente, el procedimiento de

comparación de mapas de bits se puede mejorar usando técnicas de aprendizaje de máquinas.

Formas de realización del método 500 incluyen la clasificación de la zona candidata (bloque 580). Por ejemplo, se puede utilizar un clasificador estadístico binario para clasificar la zona candidata como poseedora o no de una anotación. El clasificador puede considerar características de entrada, tales como información generada a partir de una o más de las etapas de formas de realización del método 500 antes descritas, tales como, por ejemplo, $P(s_c)$, $P(s_w)$, información de color, diferencias de píxeles en superposición de mapas de bits, las dimensiones de la zona candidata, y la posición de la zona candidata con respecto a un margen de la página impresa. Por ejemplo, el clasificador puede tomar, como entradas, x , y , w , y h , donde x es la distancia horizontal de la zona candidata con respecto al margen de la página, y es la distancia vertical de la zona candidata con respecto al margen de la página, w es la anchura de la zona candidata, y h es la altura de la zona, pudiéndose determinar todas ellas durante el fraccionamiento. En formas de realización, el clasificador puede adoptar, como entrada, un número ilimitado de tipos diferentes de información además de los correspondientes antes descritos, o en lugar de estos últimos. Adicionalmente, el clasificador se puede entrenar, usando un número ilimitado de técnicas de aprendizaje de máquinas, con el fin de mejorar su capacidad de clasificar zonas. Los ejemplos de características que se pueden usar para entrenar el clasificador incluyen color, tono, saturación, subrayado, escritura manual, fuentes, posiciones de los márgenes, ruido de píxeles, y similares.

Aunque formas de realización de la presente invención se han descrito con especificidad, la propia descripción no está destinada a limitar el alcance de esta patente. Por el contrario, los inventores han contemplado que la invención reivindicada también se podría materializar de otras maneras, para incluir diferentes etapas o características, o combinaciones de etapas o características similares a las descritas en este documento, conjuntamente con otras tecnologías.

REIVINDICACIONES

1. Método implementado por ordenador para sincronizar anotaciones entre un documento impreso y un documento electrónico, caracterizado por que comprende:

5 recibir, en un servidor, una imagen de una parte de documento impreso que tiene una parte de texto impreso, incluyendo la parte de documento impreso una anotación en la proximidad de la parte de texto impreso;

10 acceder a por lo menos una parte del documento electrónico desde un dispositivo de memoria, comprendiendo el documento electrónico una versión electrónica del documento impreso;

15 identificar una parte de texto electrónico dentro del documento electrónico, que se corresponde con la parte de texto impreso;

detectar la anotación en la parte de documento impreso; y

20 facilitar la inserción de una anotación digital en por lo menos uno de entre el documento electrónico y una copia del documento electrónico, correspondiéndose la anotación digital con la anotación detectada, siendo la anotación digital insertada en la proximidad de la parte de texto electrónico, identificada, en una posición que se corresponde por lo menos sustancialmente con una posición de la anotación detectada de la parte de documento impreso.

25 2. Método según la reivindicación 1, que comprende asimismo:

identificar una anotación digital adicional en el documento electrónico;

30 determinar que la parte de documento impreso no incluye una anotación adicional correspondiente a la anotación digital adicional;

determinar si la anotación digital adicional es una anotación migrada; y

35 eliminar la anotación digital adicional del documento electrónico si la anotación digital adicional es una anotación migrada.

3. Método según la reivindicación 2, en el que la determinación de si la anotación digital adicional es una anotación migrada comprende determinar un valor de una etiqueta asociada a la anotación digital adicional.

5

4. Método según la reivindicación 1, que comprende asimismo:

identificar un tipo de anotación correspondiente a la anotación detectada, comprendiendo el tipo de anotación por lo menos uno de entre una parte resaltada, un texto escrito a mano,
10 una parte subrayada, y un marcapáginas, y comprendiendo la facilitación de la inserción de la anotación digital en el documento electrónico la creación de una anotación digital del tipo de anotación identificado.

5. Método según la reivindicación 1, en el que la facilitación de la inserción de la anotación digital en el documento electrónico comprende insertar texto electrónico en un margen del
15 documento electrónico.

6. Método según la reivindicación 1, en el que el documento electrónico comprende un formato de documento independiente de la plataforma.

20

7. Método según la reivindicación 1, en el que la identificación de la parte de texto electrónico correspondiente comprende identificar una similitud textual entre la parte de texto impreso y la parte de texto electrónico correspondiente.

8. Método según la reivindicación 7, en el que la identificación de la similitud textual comprende:

reconocer una fracción de la parte de texto impreso llevando a cabo un procedimiento de reconocimiento óptico de caracteres (OCR) sobre la imagen de la parte de documento
30 impreso;

convertir la fracción reconocida de la parte de texto impreso en texto electrónico reconocido usando el procedimiento OCR;

identificar por lo menos una secuencia de caracteres en el texto electrónico reconocido, que incluye un error de reconocimiento;

35

crear un texto electrónico reconocido podado, comprendiendo el texto electrónico reconocido podado el texto electrónico reconocido del cual se ha eliminado dicha por lo menos una secuencia de caracteres; y

5
buscar en el documento electrónico usando por lo menos una consulta de búsqueda que comprende el texto electrónico reconocido podado.

9. Uno o más soportes legibles por ordenador, caracterizados por que comprenden
10 instrucciones ejecutables por ordenador materializadas en los mismos, para facilitar la sincronización de anotaciones entre un documento impreso y un documento electrónico, incluyendo las instrucciones una pluralidad de componentes de programa, comprendiendo la pluralidad de componentes de programa:

15 un componente de comparación que (1) recibe una imagen de una parte de documento impreso que tiene una parte de texto impreso y (2) identifica una parte de texto electrónico dentro del documento electrónico, que se corresponde con la parte de texto impreso; y

20 un componente de anotación digital que facilita la inserción de una anotación digital en el documento electrónico en la proximidad de la parte de texto electrónico correspondiente identificada.

10. Soporte según la reivindicación 9, en el que el componente de comparación utiliza un procedimiento de reconocimiento óptico de caracteres (OCR) para reconocer la parte de texto
25 impreso dentro de la parte de documento impreso, y utilizando el componente de comparación una o más consultas de búsqueda para identificar la parte de texto electrónico correspondiente.

11. Soporte según la reivindicación 10, en el que el componente de comparación utiliza un procedimiento de poda para eliminar errores de reconocimiento de la parte de texto impreso
30 reconocida.

12. Soporte según la reivindicación 9, en el que la anotación digital comprende por lo menos una de entre una anotación digital directa y una anotación digital migrada.

35 13. Soporte según la reivindicación 12, que comprende asimismo:

un componente de detección que detecta una anotación en la parte de documento impreso analizando por lo menos la imagen de la parte de documento impreso, incluyendo la parte de documento impreso la anotación en la proximidad de la parte de texto impreso.

5 14. Soporte según la reivindicación 13, en el que el componente de detección fracciona la imagen de la parte de documento impreso en una pluralidad de zonas, comprendiendo la pluralidad de zonas por lo menos una zona de texto y por lo menos una zona candidata.

10 15. Soporte según la reivindicación 14, en el que el componente de detección analiza dicha por lo menos una zona candidata usando por lo menos uno de entre un procedimiento de reconocimiento de escritura manual, un procedimiento OCR, un modelo de lenguaje estadístico, una comparación por superposición de mapas de bits, y un clasificador estadístico.

15 16. Soporte de la reivindicación 13, en el que el componente de detección detecta la anotación basándose, por lo menos en parte, en retroalimentación recibida desde por lo menos un revisor por medio de una plataforma de *crowd-sourcing*.

17. Sistema que facilita la sincronización de anotaciones entre un documento impreso y un documento electrónico, caracterizado por que comprende:

20 un servidor configurado para recibir, desde un dispositivo de formación de imágenes, una imagen de una parte de documento impreso que tiene una anotación en la proximidad de una parte de texto impreso, comprendiendo el servidor un procesador que instancia un sincronizador configurado para:

25 (a) identificar una parte de texto electrónico correspondiente en el documento electrónico, siendo la parte de texto electrónico correspondiente textualmente similar a la parte de texto impreso,

30 (b) detectar la anotación en la parte de documento impreso, y

(c) facilitar la inserción de una anotación digital en el documento electrónico, correspondiendo la anotación digital a la anotación detectada.

35 18. Sistema según la reivindicación 17, en el que el sincronizador está configurado para detectar la anotación utilizando por lo menos uno de entre un procedimiento de reconocimiento

de escritura manual, un procedimiento OCR, un modelo de lenguaje estadístico, una comparación por superposición de mapas de bits, y un clasificador estadístico.

5 19. Sistema según la reivindicación 17, en el que el sincronizador está configurado para facilitar la inserción de la anotación digital en el documento electrónico (1) asociando al documento electrónico un archivo de metadatos que contiene la anotación, y (2) insertando en el documento electrónico un puntero al archivo de metadatos.

10 20. Sistema según la reivindicación 17, en el que el dispositivo de formación de imágenes comprende por lo menos uno de entre una cámara dispuesta en un dispositivo de lectura y un dispositivo de escaneado industrial.

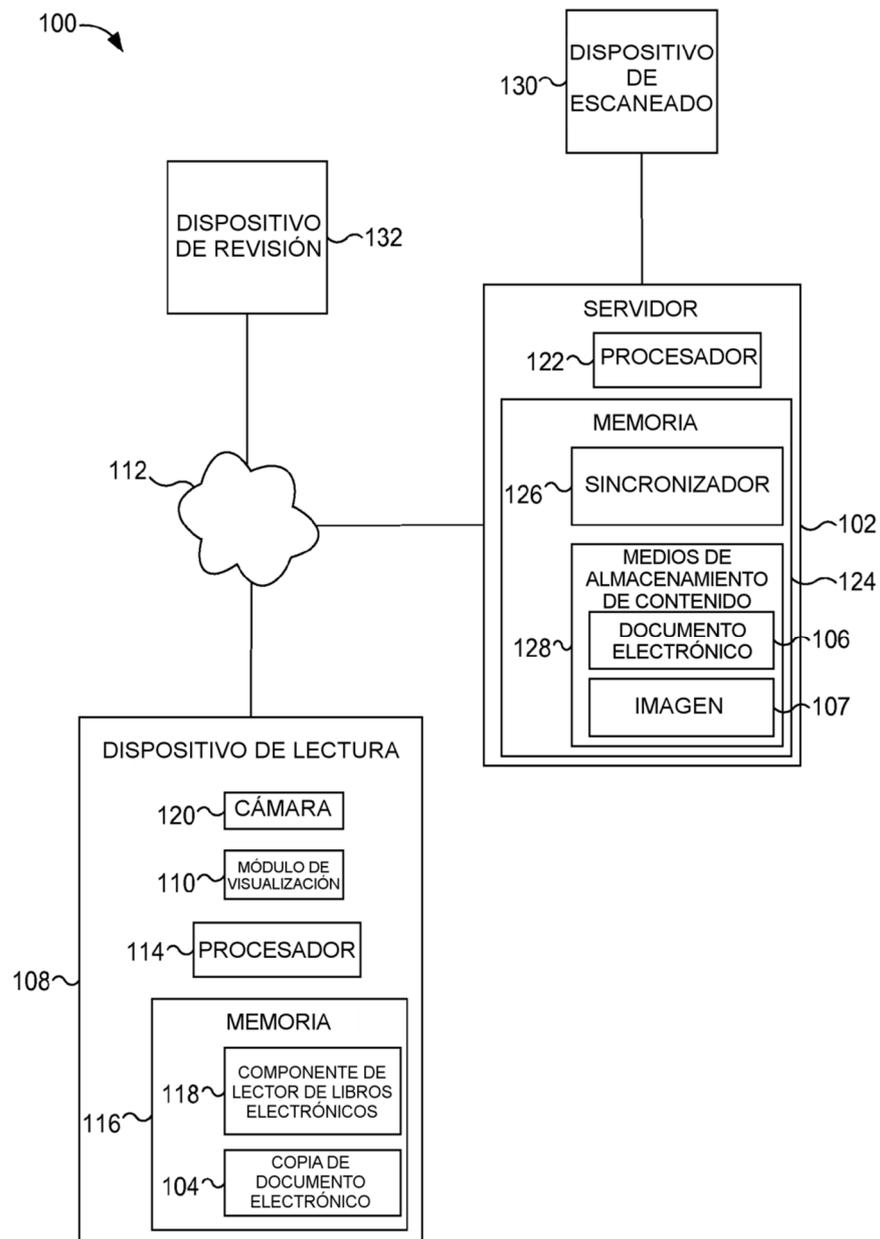


FIG. 1

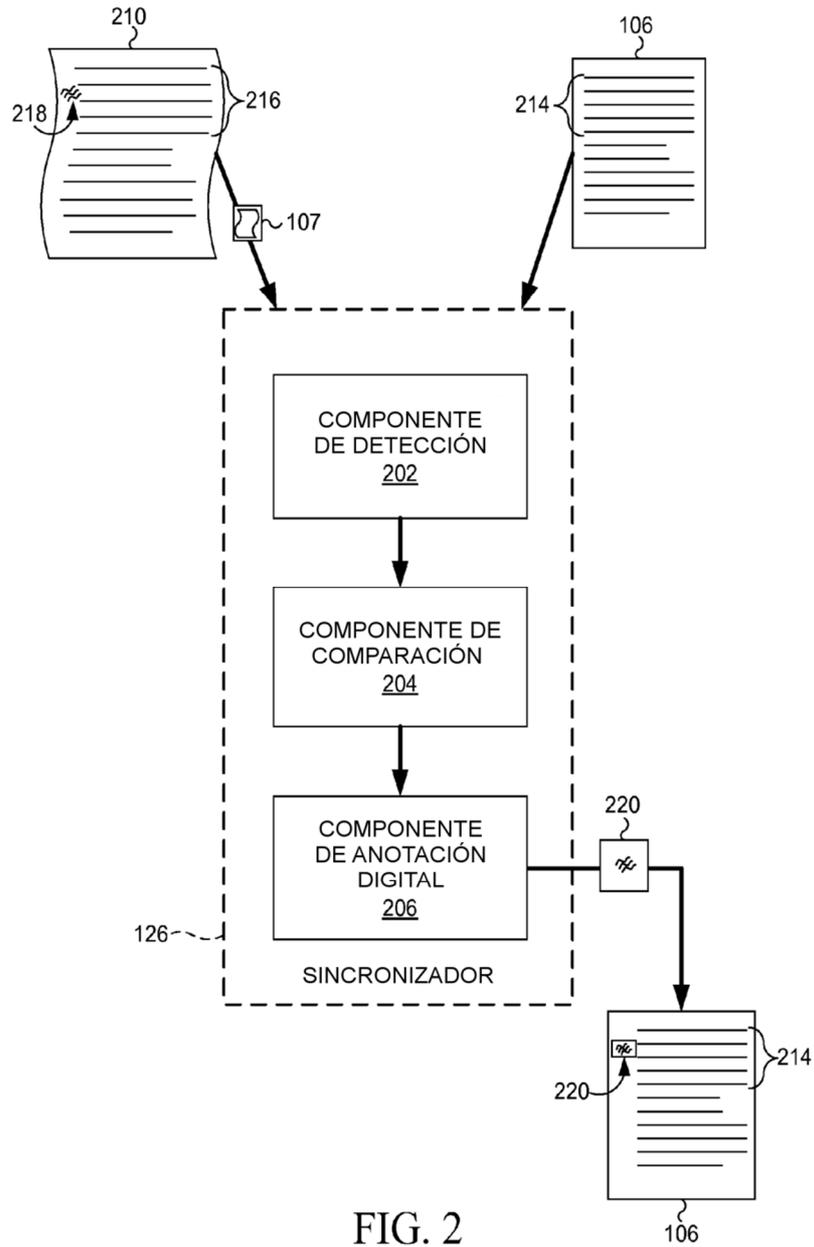


FIG. 2

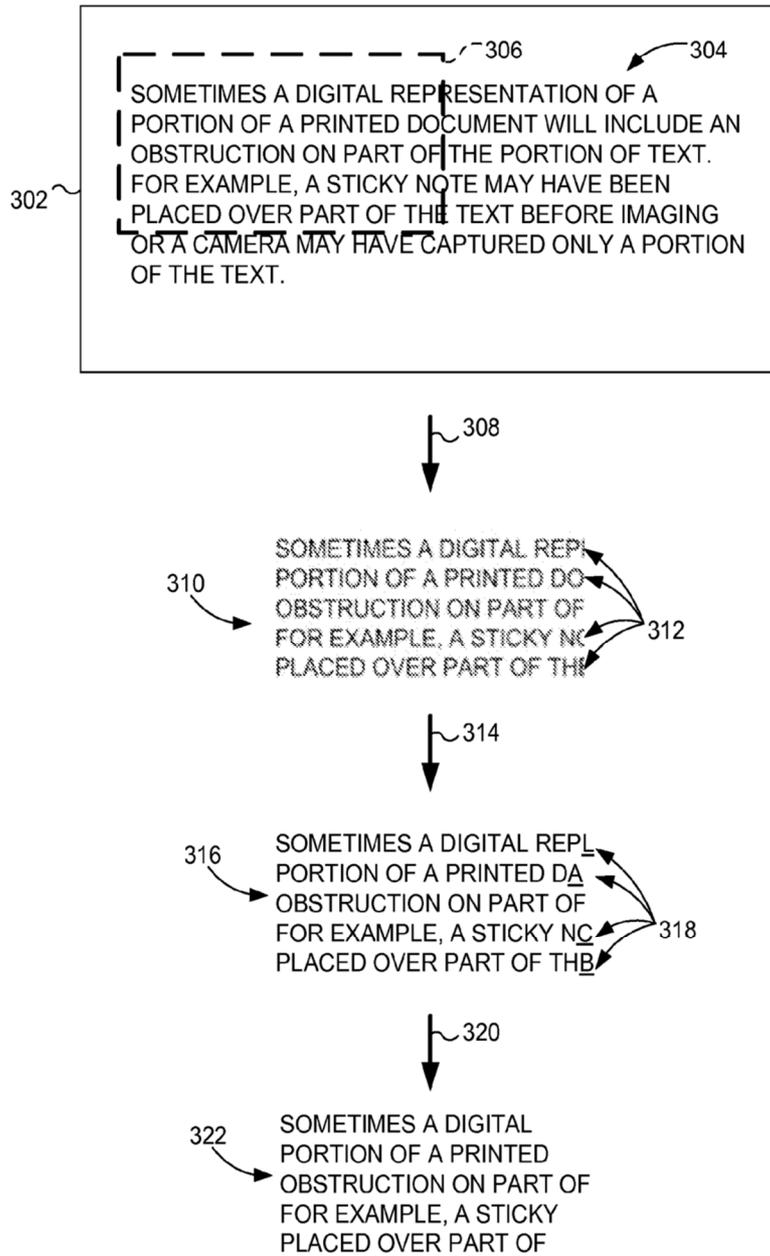


FIG. 3

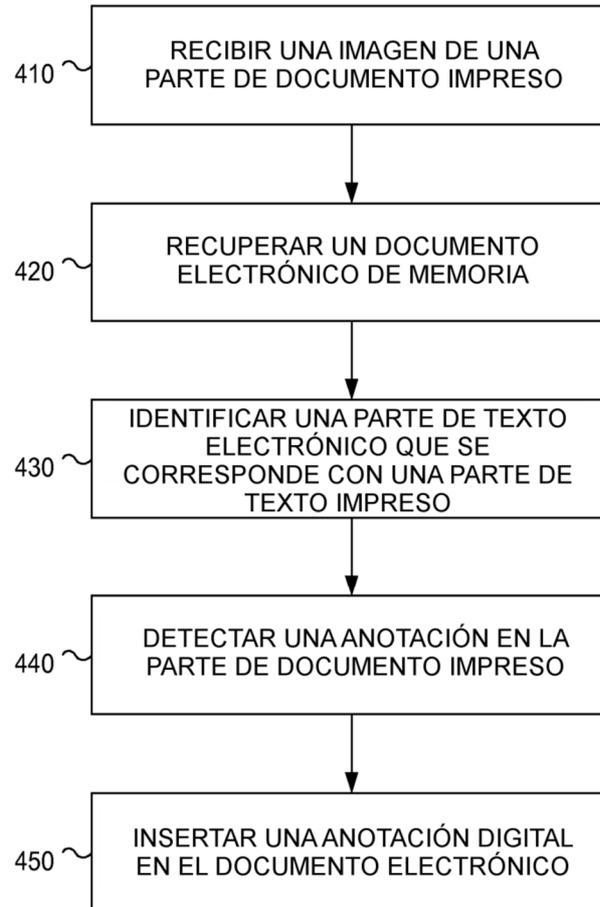
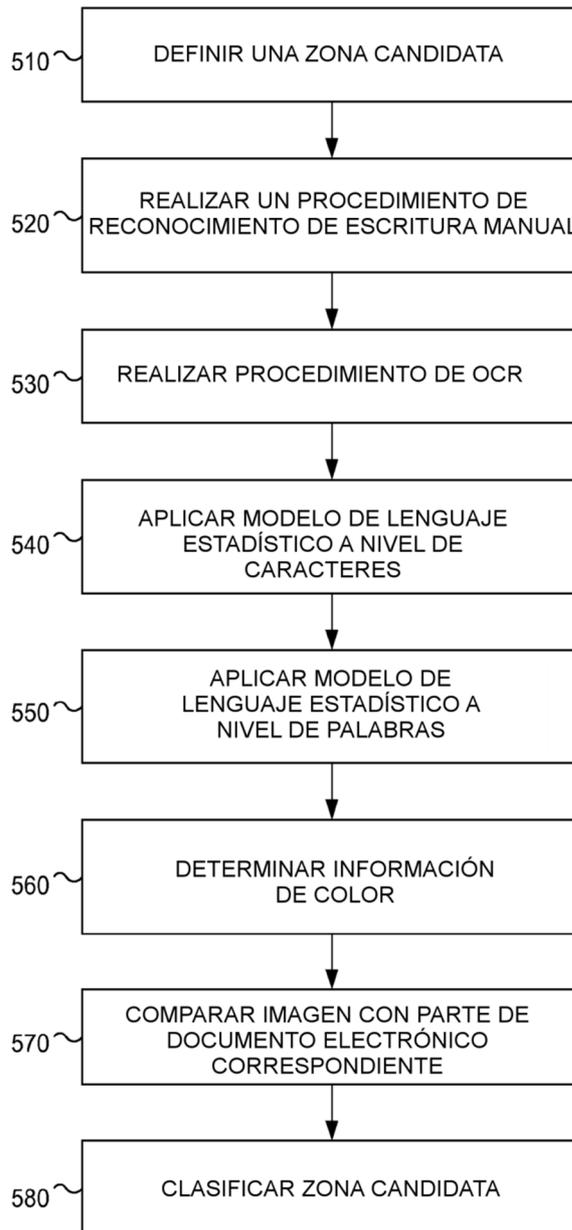


FIG. 4

400 ↗



500 ↗

FIG. 5