

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 985 687**

51 Int. Cl.:

B42F 21/06 (2006.01)

G06F 3/00 (2006.01)

G06F 7/00 (2006.01)

G06F 3/12 (2006.01)

G06F 16/31 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.08.2020 PCT/US2020/048629**

87 Fecha y número de publicación internacional: **04.03.2021 WO21041983**

96 Fecha de presentación y número de la solicitud europea: **28.08.2020 E 20857785 (8)**

97 Fecha y número de publicación de la concesión europea: **17.04.2024 EP 4004690**

54 Título: **Métodos para indexar y recuperar texto**

30 Prioridad:

30.08.2019 US 201962894386 P

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

07.11.2024

73 Titular/es:

**SHOEIBI, LISA (100.0%)
4500 Williams Dr. Ste 212 145
Georgetown TX 78633, US**

72 Inventor/es:

SHOEIBI, LISA

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 985 687 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

Métodos para indexar y recuperar texto

Campo de la Invención

5 La presente divulgación se refiere en general a sistemas de indexación. Más específicamente, la presente divulgación describe un método para indexar y recuperar texto.

Antecedentes de la invención

10 La recuperación de información es la actividad de obtener recursos de sistema de información que son pertinentes para una necesidad de información de una recopilación de aquellos recursos. Las búsquedas pueden ser con base en texto completo u otra indexación basada en contenido. La recuperación de información incluye la ciencia de buscar información en un documento, buscar documentos en sí mismos y también buscar los metadatos que describen los datos, y las bases de datos de textos, imágenes o sonidos. La presente indexación es el acto de describir o clasificar un documento por términos de índice u otros símbolos a fin de indicar de qué se trata el documento, resumir su contenido o incrementar su capacidad de búsqueda. En otras palabras, se trata de identificar y describir el tema de los documentos. Los índices se pueden construir, por separado, en tres niveles distintos: términos en un documento tal como un libro; objetos en una colección tal como una biblioteca; y documentos (tal como libros y artículos) dentro de un campo de conocimiento.

Breve descripción de las figuras

Algunas de las realizaciones se describirán en detalle, con referencia a las siguientes figuras, en donde designaciones similares indican miembros similares, en donde:

La figura 1 representa una primera hoja.

20 La figura 2 representa una segunda hoja.

La figura 3 representa una tercera hoja, que no forma parte de las reivindicaciones.

La figura 4 representa una cuarta hoja, que no se reivindica.

La figura 5 representa una colección física de texto, así como identificadores de texto e identificadores de página fijados a la misma.

25 La figura 6 representa los pasos de proceso de un método para indexar y recuperar texto, de acuerdo con la invención.

La figura 7 representa los pasos de proceso de un método para indexar y recuperar texto, de acuerdo con ciertas realizaciones.

La figura 8 representa los pasos de proceso de un método para indexar y recuperar texto, de acuerdo con aún otras realizaciones.

30 Salvo que se indique específicamente lo contrario, los artículos representados en las figuras no necesariamente se dibujan a escala.

Descripción detallada de la invención

35 Es importante señalar que cada término utilizado en el presente documento se refiere a lo que un experto en la técnica entendería que este término significa con base en el uso contextual de este término en la presente. En la medida en que el significado de un término usado en la presente, como se entiende por el experto en la técnica con base en el uso contextual de este término, difiera de alguna manera de cualquier definición particular del diccionario de este término, se propone que el significado del término como se entiende por el experto en la técnica debe prevalecer.

40 Además, es importante tener en cuenta que, como se usa en la presente, "uno" y "una" en general denota "al menos uno", pero no excluye una pluralidad salvo que el uso contextual dicte lo contrario. Cuando se utiliza en la presente para unirse a una lista de artículos, "o" denota "al menos uno de los artículos", pero no excluye una pluralidad de artículos de la lista. Finalmente, cuando se utiliza en la presente para unirse a una lista de artículos, "y" denota "todos los artículos de la lista".

Varias técnicas anteriores cercanas están disponibles para métodos para indexar y recuperar texto. Por ejemplo, US9987877B1 emitido para Dunn divulgó una herramienta y métodos de indexación visual. Dunn divulga un sistema y

método para indexar y usar de manera efectiva un extenso libro de referencia, tal como la Biblia, por tema, tal que se pueda acceder de manera conveniente y rápidamente a la información apropiada con relación a un tema deseado de manera oportuna y útil. US20080263032A1 emitido para Vailaya, Wu y Rathi divulga el procesamiento y la búsqueda de documentos no estructurados y semiestructurados. Este documento divulga un método para analizar e indexar un documento no estructurado o semiestructurado de acuerdo con una realización que incluye recibir un documento no estructurado o semiestructurado; convertir el documento en uno o más flujos de texto; analizar el uno o más flujos de texto para identificar los contenidos textuales del documento; analizar el uno o más flujos de texto para identificar secciones lógicas del documento; asociar los contenidos textuales con las secciones lógicas; indexar el contenido textual y su asociación con las secciones lógicas; y guardar un resultado de la indexación en un dispositivo de almacenamiento de datos. US20080027933A1 se emite para Hussam. El documento divulga el sistema y método para la ubicación, comprensión y asimilación de documentos digitales a través de indicios abstractos.

La siguiente descripción detallada se refiere a las figuras anexas. Siempre que sea posible, se utilizan los mismos números de referencia en las figuras y la siguiente descripción para referirse a los mismos elementos o similares. En tanto que se pueden describir muchas realizaciones de la divulgación, son posibles modificaciones, adaptaciones, y otras implementaciones dentro del alcance de las reivindicaciones anexas. Por ejemplo, se pueden realizar sustituciones, adiciones, o modificaciones a los elementos ilustrados en las figuras, y los métodos descritos en la presente se pueden modificar por la sustitución, reordenación, o adición de etapas a los métodos divulgados. Sin embargo, el alcance de la invención se define por las reivindicaciones anexas. La presente divulgación contiene encabezamientos. Se debe entender que estos encabezamientos se utilizan como referencias y no se van a interpretar como limitantes de la presente materia divulgada bajo el encabezamiento.

Otras ventajas técnicas pueden llegar a ser fácilmente evidentes para un experto en la técnica después de la revisión de las siguientes figuras y descripción. Se debe entender desde el principio que, aunque se ilustran realizaciones de ejemplo en las figuras y se describen a continuación, los principios de la presente divulgación se pueden implementar usando cualquier cantidad de técnicas, ya sean conocidas actualmente o no.

La recuperación de información es la actividad de obtener recursos de sistema de información que son pertinentes para una necesidad de información de una recopilación de aquellos recursos. Las búsquedas pueden ser con base en texto completo u otra indexación basada en contenido. La recuperación de información incluye la ciencia de buscar información en un documento, buscar documentos en sí mismos y también buscar los metadatos que describen los datos, y las bases de datos de textos, imágenes o sonidos. La presente indexación es el acto de describir o clasificar un documento por términos de índice u otros símbolos a fin de indicar de qué se trata el documento, resumir su contenido o incrementar su capacidad de búsqueda. En otras palabras, se trata de identificar y describir el tema de los documentos.

Los índices se pueden construir, por separado, en tres niveles distintos: términos en un documento tal como un libro; objetos en una colección tal como una biblioteca; y documentos (tal como libros y artículos) dentro de un campo de conocimiento. Las realizaciones divulgadas en la presente buscan proporcionar métodos para la indexación de texto que se incluye en una colección física de texto. Como se utiliza en la presente, "colección física de texto" se refiere a libros y publicaciones periódicas (por ejemplo, periódicos, revistas, boletines, revistas, y publicaciones en serie similares que aparecen en una nueva edición con regularidad). Las realizaciones de la presente solicitud buscan divulgar métodos para recuperar el texto indexado.

El método reivindicado incluye el uso de una primera hoja 100 y una segunda hoja 200. El uso de una tercera hoja 300 y una cuarta hoja 440 es posible pero no es parte de las reivindicaciones. La primera hoja 100, la segunda hoja 200, la tercera hoja 300, y/o la cuarta hoja 400 ("las hojas") cada una puede incluir una pluralidad de papel impreso, plástico, vinilo, u otro material que tiene adhesivo sensible a la presión en un lado (por ejemplo, pegatinas, etiquetas adhesivas, avisos adhesivos, etc.). Por ejemplo, el adhesivo sensible a la presión se puede utilizar para la fijación temporal o permanente. Cada una de las hojas puede ser o incluir cualquier color, forma, y/o tamaño para facilitar una o más realizaciones divulgadas en la presente. Como se representa en la figura 1, la primera hoja 100 incluye preferentemente una pluralidad de identificadores de texto (por ejemplo, identificadores de texto 110 e identificadores de texto 115) en el adhesivo sensible a la presión en un lado.

Por ejemplo, cada identificador de texto 110 incluye una enumeración 120 y cada identificador de texto 115 incluye una enumeración 125. Cada conjunto de identificadores de texto incluye preferentemente una iconografía específica (por ejemplo, una forma y/o color específico). Por ejemplo, cada identificador de texto 110 incluye una iconografía 130, que es de forma hexagonal. De manera similar, cada identificador de texto 115 incluye una iconografía 135, que es de forma triangular. La figura 2 representa la segunda hoja, en general 200, de acuerdo con ciertas realizaciones. La segunda hoja 200 incluye preferentemente una pluralidad de identificadores de página (por ejemplo, identificadores de página 210 e identificadores de página 215). Por ejemplo, cada identificador de página 210 incluye una enumeración 230 y cada identificador de página 215 incluye una enumeración 235. Cada conjunto de identificadores de página incluye preferentemente una iconografía específica (por ejemplo, una forma y/o color específico).

Por ejemplo, cada identificador de página 210 incluye una iconografía 130 y cada identificador de página 215 incluye una iconografía 135. En otras palabras, los identificadores de texto y los identificadores de página que tienen la misma

iconografía y enumeraciones se configuran para usarse conjuntamente. La figura 3 representa un ejemplo de una tercera hoja, en general 300, que no forma parte de las reivindicaciones. La tercera hoja 300 puede incluir una lista 305, que comprende una tabla de celdas 306 arregladas en filas (por ejemplo, filas 315) y columnas (por ejemplo, columnas 320 y 325). En otros ejemplos, una fila 315 incluye una o más filas subordinadas. Algunas filas de la lista se utilizan para almacenar resúmenes de texto (por ejemplo, resumen de texto 310 y resumen de texto 330). Algunas filas incluyen las enumeraciones 335. En realizaciones preferidas, cada fila se asocia con una iconografía (por ejemplo, la iconografía 110 y la iconografía 115). Aquí, cada iconografía (por ejemplo, la iconografía 110) se asocia con un artículo de indexación 316 (por ejemplo, miedo, promesas, curación, ángeles, fuerza). En realizaciones preferidas, cada fila se enumera (es decir, incluye una enumeración 330) para coincidir con la enumeración de un identificador de texto e identificador de página particulares. Por ejemplo, el resumen de texto 310 ("RESUMEN DE TEXTO A") coincide con las enumeraciones 120 y 230 del identificador de texto 110 y la página 210.

La figura 4 representa una cuarta hoja, en general 400, de acuerdo con ciertas realizaciones. La cuarta hoja 400 incluye preferentemente las mismas iconografías que la primera hoja 100 y la segunda hoja 200 (por ejemplo, la iconografía 110 y la iconografía 115). Las iconografías incluidas en la cuarta hoja 400 se pueden fijar a la lista 305 para asociar el artículo de indexación 316 con una iconografía particular (por ejemplo, la iconografía 110). Por ejemplo, la iconografía 115 se fija preferentemente a la lista 305 para asociar de esta manera la iconografía con el artículo de indexación 317. Por lo tanto, los identificadores de texto (por ejemplo, el identificador de texto 110) y los identificadores de página (por ejemplo, el identificador de página 210) que incluyen cada uno, por ejemplo, la iconografía 110 se asocian con el artículo de indexación 316. Los identificadores de texto y los identificadores de página que tienen las mismas enumeraciones están con relación.

Las figuras 5-8 se utilizan en la presente para representar los pasos de proceso de un método para permitir la indexación y recuperación de texto, de acuerdo con otras realizaciones. Sin duda, los pasos del proceso descritos en la presente se pueden modificar al sustituir, reordenar, o adicionar un paso a los métodos divulgados. Como se utiliza en la presente, los identificadores de texto 510 y 505 son identificadores de texto que realizan las mismas funciones que los identificadores de texto 110 y 115 (por ejemplo, cada identificador de texto incluye una iconografía y enumeración). Los identificadores de página 515 y 520 son identificadores de página que realizan las mismas funciones e incluyen los mismos componentes que los identificadores de página 210 y 215 (por ejemplo, cada identificador de página incluye una iconografía y enumeración). El identificador de texto 505 y el identificador de página 515 incluyen cada uno la misma iconografía (por ejemplo, un corazón) y enumeración (por ejemplo, "1"). De manera similar, el identificador de texto 510 y el identificador de página 520 incluyen cada uno la misma iconografía (por ejemplo, un hexágono) y enumeración (por ejemplo, "4"). En el paso 600, un primer identificador de texto (por ejemplo, el identificador de texto 505) se fija adyacente a un primer cuerpo de texto (por ejemplo, texto 525), el primer cuerpo de texto se incluye en una colección física de texto 500, y el primer identificador de texto comprende una primera iconografía (es decir, la forma de corazón). En el paso 605, se fija un primer identificador de página adyacente a una página 535 incluida en la colección física de texto 535, la página 535 incluye el primer cuerpo de texto, y el primer identificador de página comprende la primera iconografía. En el paso 610, se modifica una lista 305 para incluir un primer resumen del primer cuerpo de texto (por ejemplo, similar al resumen de texto 330), la lista 305 es una tabla de celdas arregladas en filas (por ejemplo, fila 315) y columnas (por ejemplo, columnas 320 y 325). Aquí, cada una de las filas se asocia con una de una pluralidad de iconografías (por ejemplo, la iconografía 110 y la iconografía 115). La pluralidad de iconografías incluye la primera iconografía y cada una se asocia con un artículo de indexación (por ejemplo, el artículo de indexación 316 y el artículo de indexación 317).

En el paso 615, un segundo identificador de texto (por ejemplo, el identificador de texto 510) se fija adyacente a un segundo cuerpo de texto (por ejemplo, texto 530), el segundo cuerpo de texto incluido en la colección física de texto, y el segundo identificador de texto incluye una segunda iconografía (por ejemplo, la forma de hexágono). En el paso 620, se fija un segundo identificador de página adyacente a una segunda página (por ejemplo, la página 535 u otra página) incluida en la colección física de texto, la segunda página incluye el segundo cuerpo de texto, el segundo identificador de página incluye la segunda iconografía, y la segunda iconografía no es la primera iconografía. En realizaciones donde la misma iconografía se comparte por la primera iconografía y la segunda iconografía, cada una incluiría una enumeración diferente. En el paso 625, la lista 305 se modifica para incluir un segundo resumen del segundo cuerpo de texto (por ejemplo, el resumen de texto 310), la pluralidad de iconografías comprende la segunda iconografía. En el paso 630, se forma una primera hoja (por ejemplo, la hoja 100) que incluye una pluralidad de los primeros identificadores de texto (por ejemplo, los identificadores de texto 110) y una pluralidad de los segundos identificadores de texto (por ejemplo, los identificadores de texto 115). En el paso 635, cada uno de la pluralidad de los primeros identificadores de texto se enumeran (por ejemplo, incluyen las enumeraciones 120) para corresponder a una de las filas 315 de la lista 305.

En el paso 640, cada uno de la pluralidad de los segundos identificadores de texto (por ejemplo, los identificadores de texto 115) se enumeran (por ejemplo, incluyen las enumeraciones 125) para corresponder a una de las filas 315 de la lista 305. En el paso 645, se forma una segunda hoja (por ejemplo, la hoja 200) que incluye una pluralidad de los primeros identificadores de página (por ejemplo, los identificadores de página 210) y una pluralidad de los segundos identificadores de página (por ejemplo, los identificadores de página 215). En el paso 650, cada uno de la pluralidad de los primeros identificadores de página se enumeran (es decir, cada uno incluye una enumeración 230) para corresponder a una de las filas 315 de la lista 305. En el paso 655, cada uno de la pluralidad de los segundos identificadores de página se enumeran (es decir, cada uno incluye una enumeración 235) para corresponder a una de las filas 315 de la lista 305.

5 En algunas realizaciones, en el paso 700, una primera fila (por ejemplo, la fila que incluye el resumen de texto 330) incluida en las filas se modifica para lograr el paso 610. Aquí, la primera fila se asocia con la primera iconografía y la primera fila incluye una primera enumeración (por ejemplo, "1") que corresponde al primer identificador de texto y al primer identificador de página, como se representa en la figura 7A. Para recuperar el texto asociado con un resumen de texto, se ubica el identificador de página que comparte la misma iconografía y enumeración que el resumen de texto. A continuación, el texto asociado con el resumen de texto se recupera cuando se ubica el identificador de texto que comparte la misma iconografía y enumeración que el identificador de página.

10 En ciertas realizaciones, en el paso 705, una segunda fila (por ejemplo, la fila que incluye el resumen de texto 310) incluida en las filas 315 se modifica para lograr el paso 625, la segunda fila asociada con la segunda iconografía, la segunda fila comprende una segunda enumeración (por ejemplo, "4") que corresponde al segundo identificador de texto y al segundo identificador de página, como se representa en la figura 7B. Sin duda, cada una de las diferentes iconografías puede incluir la misma enumeración, puesto que cada iconografía se asocia con un artículo de indexación particular. En otras realizaciones, la colección física de texto 500 se forma para ser un libro o una publicación periódica en el paso 805. En el paso 810, la colección física de texto 500 se forma para incluir al menos la primera hoja 100, la segunda hoja 200 y la
15 tercera hoja 300 (es decir, la lista 305). En otras palabras, la colección física de texto 500 puede incluir la primera hoja 100, la segunda hoja 200 y la tercera hoja 300 para facilitar la indexación y recuperación de la información incluida en la misma.

REIVINDICACIONES

1. Un método para indexar y recuperar texto, que comprende los pasos de:

5 fijar un primer identificador de texto (110, 510) adyacente a un primer cuerpo de texto (525), el primer cuerpo de texto (525) se incluye en una colección física de texto (500), el primer identificador de texto (110, 510) comprende una primera iconografía (130, 135);

fijar un primer identificador de página (210, 215) adyacente a una página (535) incluida en la colección física de texto (500), la página (535) incluye el cuerpo de texto particular, el primer identificador de página (110, 510) comprende la primera iconografía (130);

10 modificar una lista (305) para comprender un primer resumen (310) del primer cuerpo de texto (525), la lista (305) comprende una tabla de celdas arregladas en filas y columnas, cada una de las filas se asocia con una de una pluralidad de iconografías (130, 135, ...), la pluralidad de iconografías (130, 135, ...) comprenden la primera iconografía (130), cada iconografía (130, 135, ...) se asocia con un artículo de indexación (316, 317);

15 fijar un segundo identificador de texto (115, 510) adyacente a un segundo cuerpo de texto (530), el segundo cuerpo de texto (530) incluido en la colección física de texto (500), el segundo identificador de texto (115, 510) comprende una segunda iconografía; fijar un segundo identificador de página adyacente a una segunda página incluida en la colección física de texto (500), la segunda página incluye el segundo cuerpo de texto, el segundo identificador de página comprende la segunda iconografía, la segunda iconografía no es la primera iconografía (130, 135);

y modificar la lista (305) para comprender un segundo resumen (310) del segundo cuerpo de texto, la pluralidad de iconografías (130, 135,...) comprenden la segunda iconografía;

20 formar una primera hoja (100), la primera hoja (100) comprende una pluralidad de los primeros identificadores de texto (110, 510) y una pluralidad de los segundos identificadores de texto;

enumerar la pluralidad de los primeros identificadores de texto (110, 510) para cada uno corresponde a una de las filas de la lista (305);

25 enumerar la pluralidad de los segundos identificadores de texto para cada uno corresponde a una de las filas de la lista (305);

formar una segunda hoja (200), la segunda hoja (200) comprende una pluralidad de los primeros identificadores de página (110, 510) y una pluralidad de los segundos identificadores de página;

enumerar la pluralidad de los primeros identificadores de página (110, 510) para cada uno corresponde a una de las filas de la lista (305); y

30 enumerar la pluralidad de los segundos identificadores de página para cada uno corresponde a una de las filas de la lista (305).

2. El método de la reivindicación 1, que comprende además formar la colección física de texto (500), la colección física de texto (500) es un libro o una publicación periódica.

35 3. El método de la reivindicación 2, en donde formar la colección física de texto (500) comprende formar la colección física de texto (500) para incluir la primera hoja (100), la segunda hoja (200) y la lista (305).

4. El método de acuerdo con la reivindicación 3, en donde modificar la lista (305) para comprender el primer resumen (330) del primer cuerpo de texto comprende:

40 modificar una primera fila incluida en las filas, la primera fila asociada con la primera iconografía (130,135), la primera fila comprende una primera enumeración (120) que corresponde al primer identificador de texto (110,510) y al primer identificador de página (210, 215).

5. El método de la reivindicación 4, en donde modificar la lista (305) para comprender el segundo resumen del segundo cuerpo de texto comprende:

45 modificar una segunda fila incluida en las filas (315), la segunda fila asociada con la segunda iconografía, la segunda fila comprende una segunda enumeración (235) que corresponde al segundo identificador de texto (510) y al segundo identificador de página.

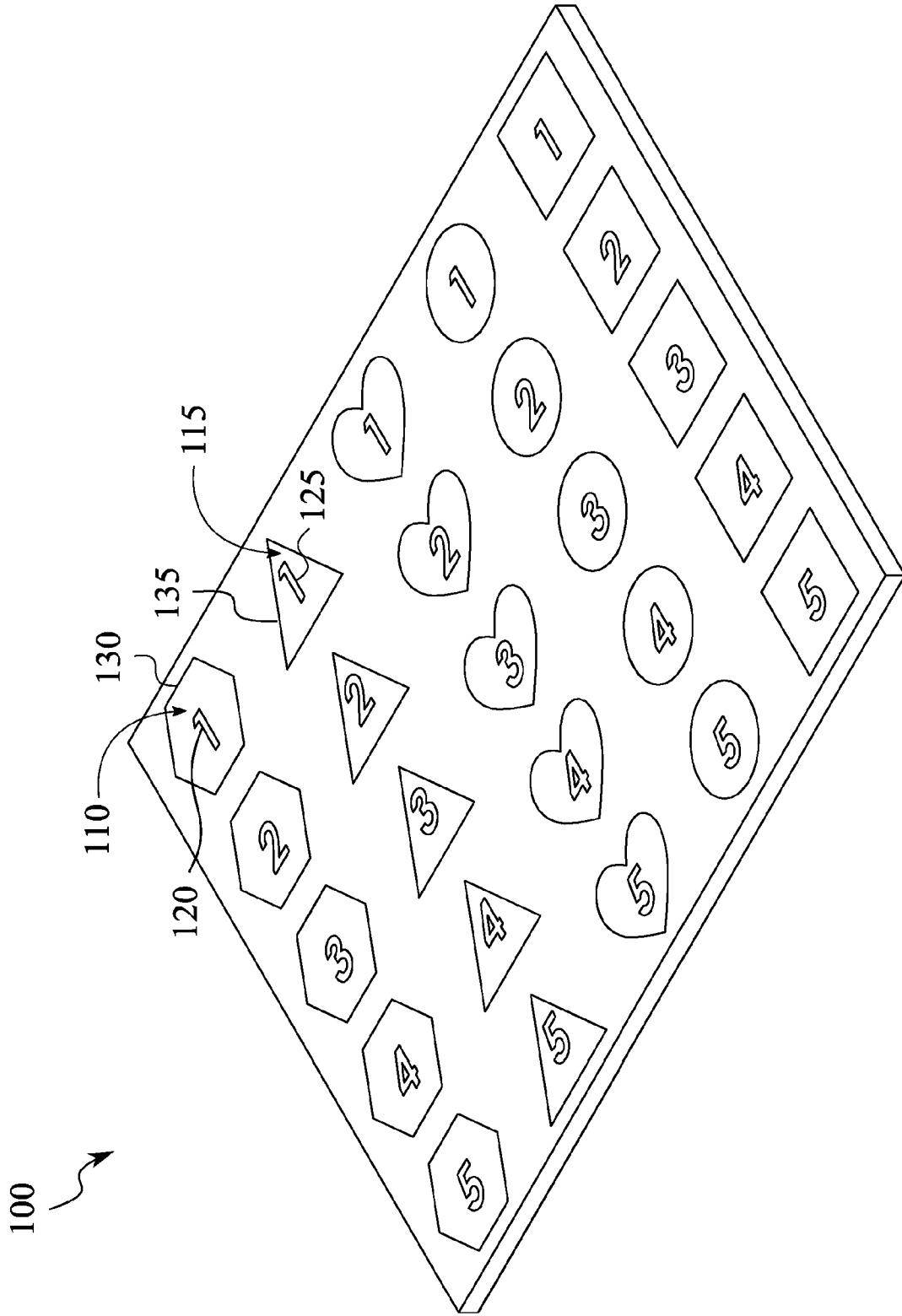


FIGURA 1

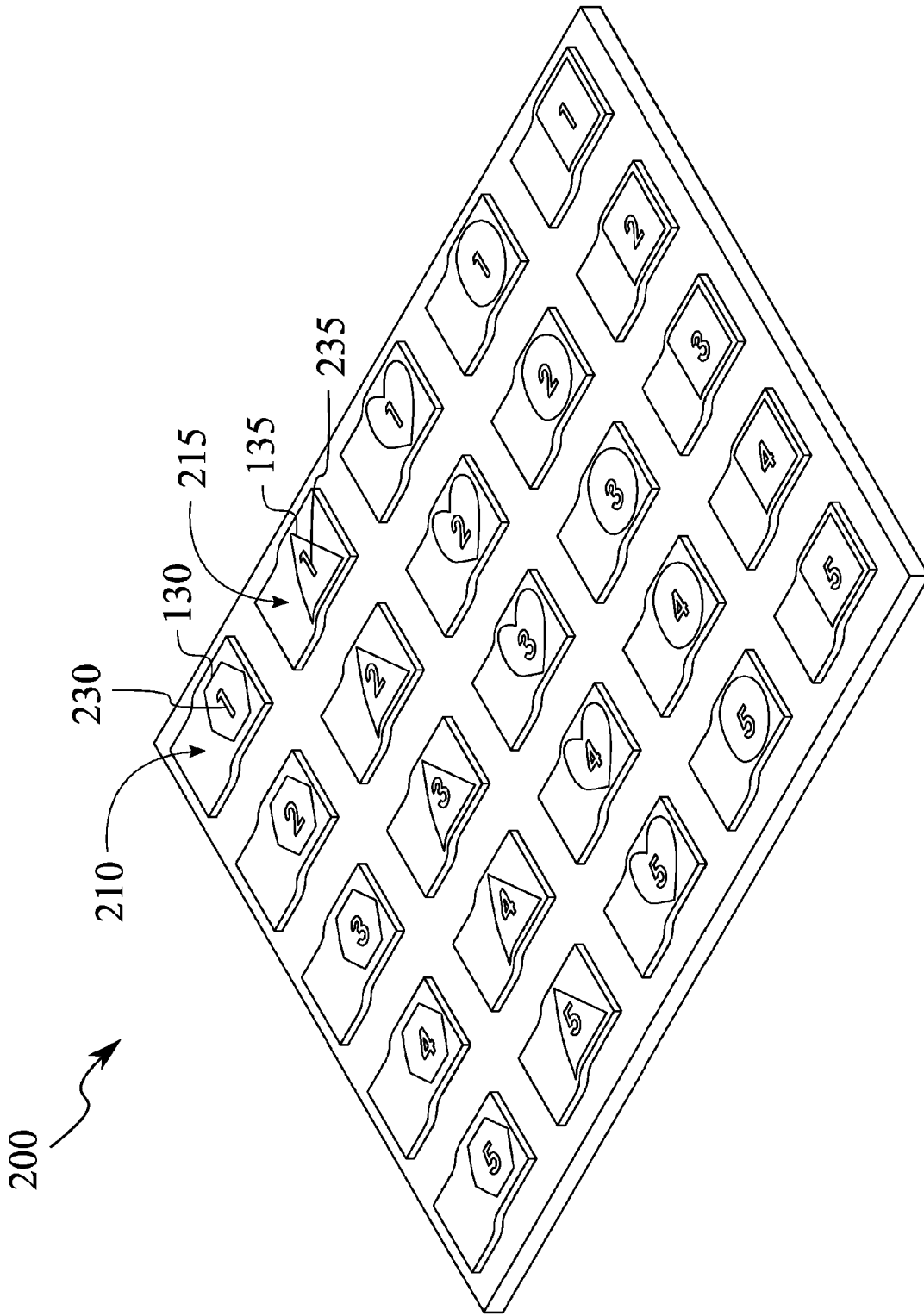


FIGURA 2

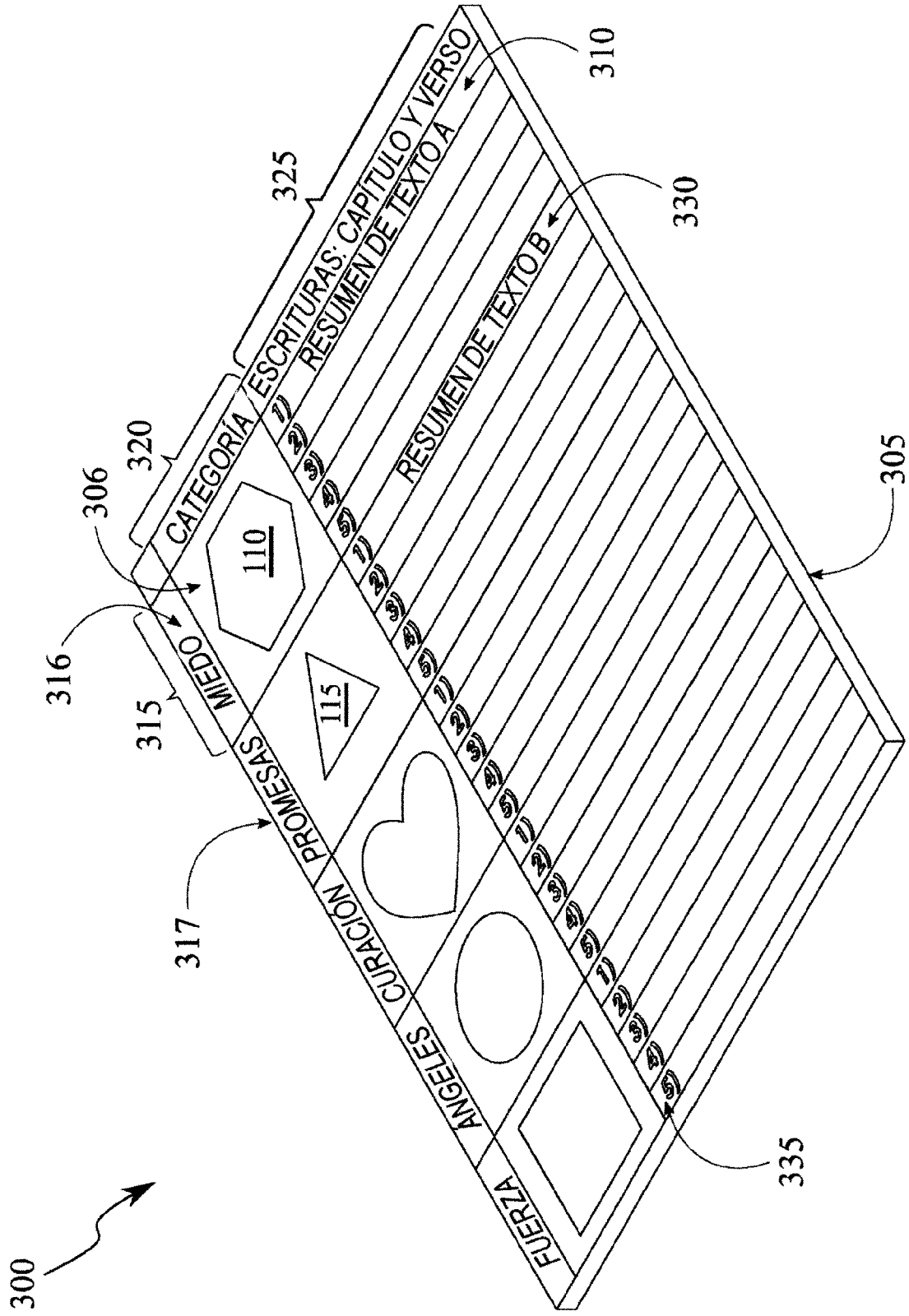


FIGURA 3

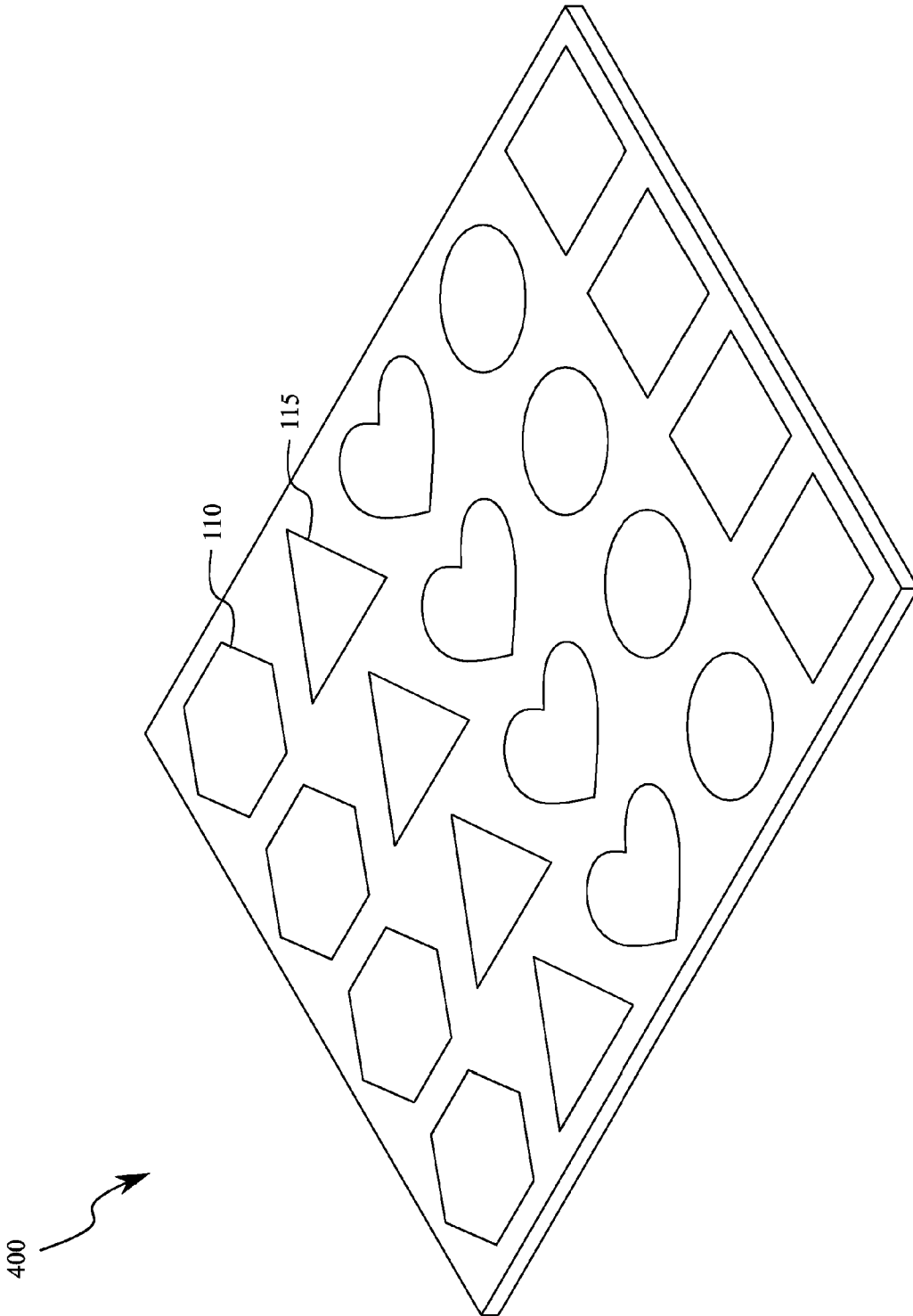


FIGURA 4

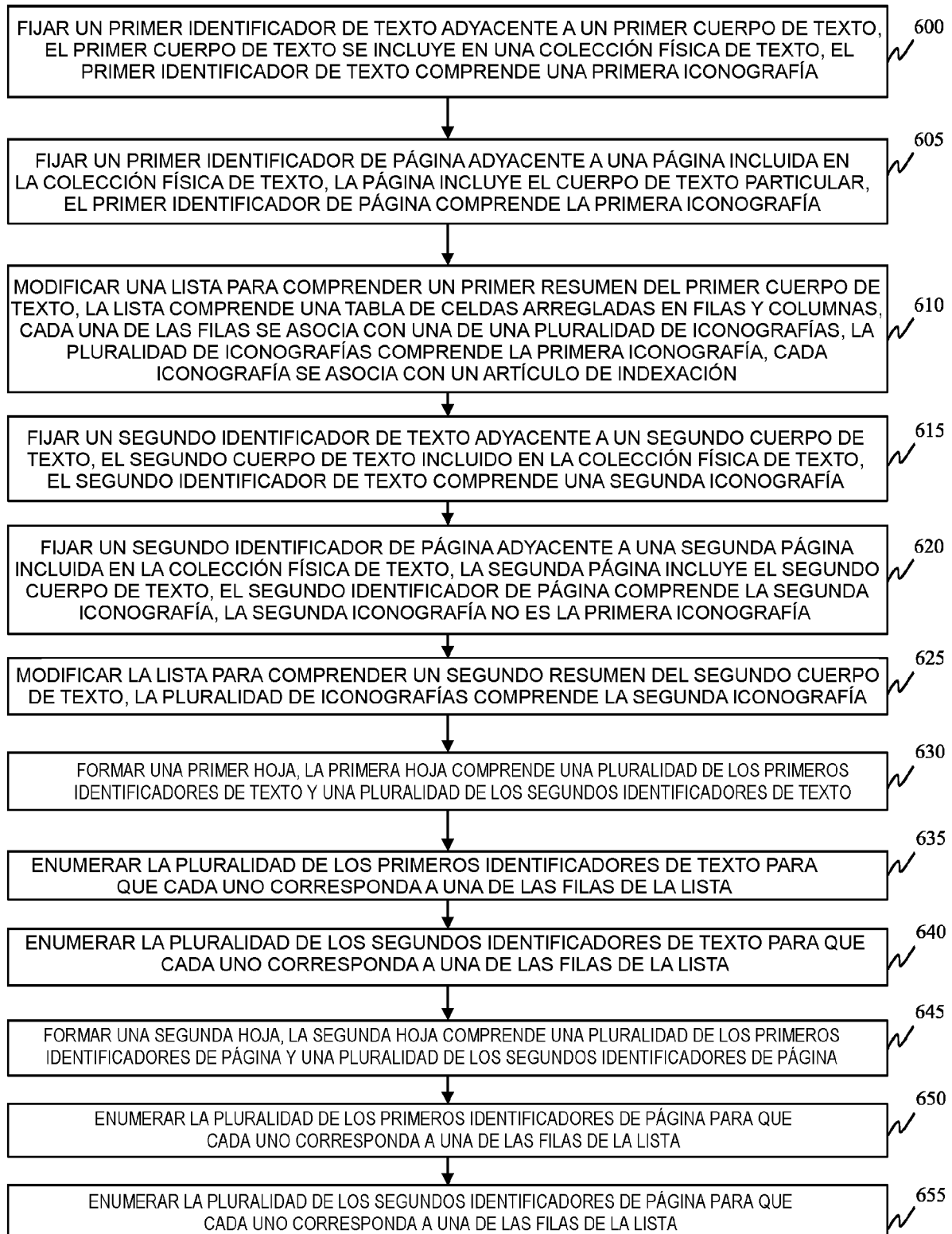


FIGURA 6

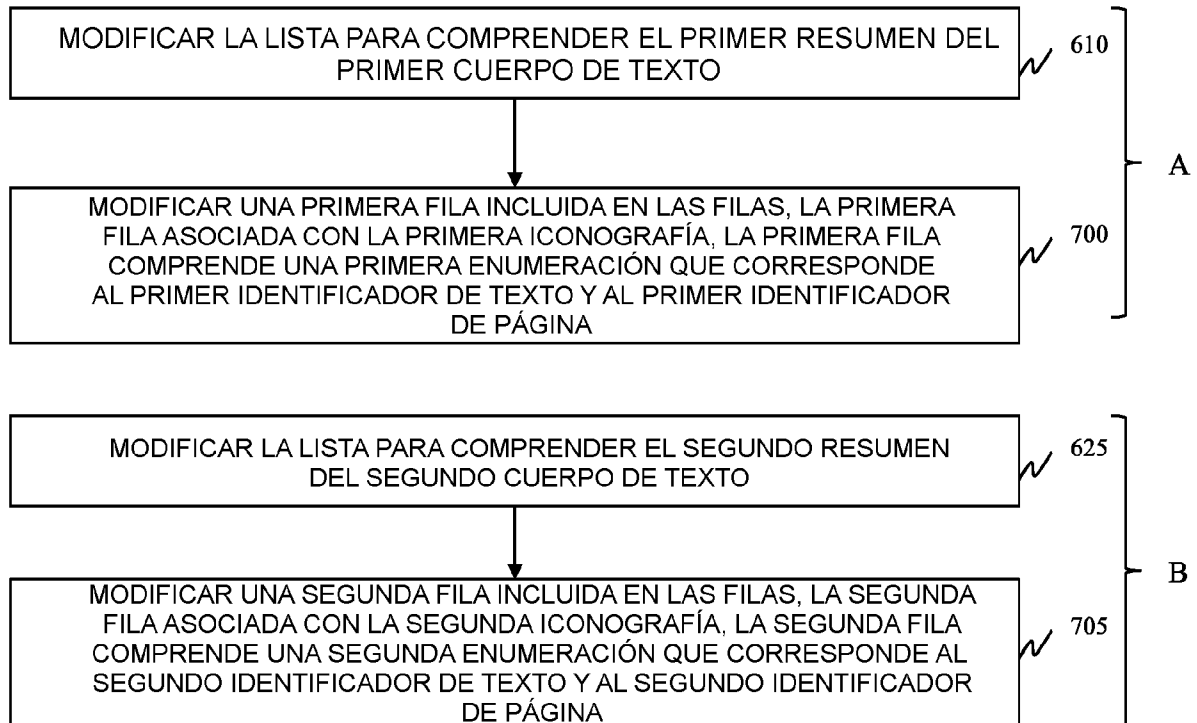


FIGURA 7

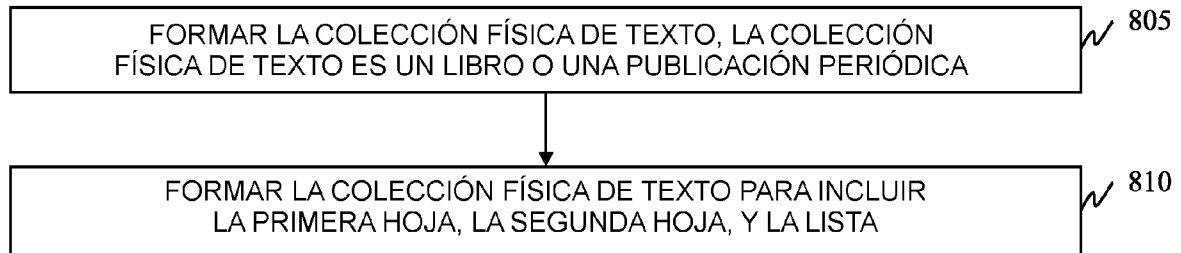


FIGURA 8