

**CAPÍTULO
REIVINDICATORIO
QUE CONTIENE LAS
MODIFICACIONES
HECHAS EN FASE
INTERNACIONAL
BAJO EL REPORTE
PRELIMINAR
INTERNACIONAL**

REIVINDICACIONES

1. Un método de caracterización automática de documento, el cual recibe un documento dado de entrada y sin estructurar, específicamente documentos legales, y da como resultado la asignación automática de una o más clases o categorías de documentos, con los cuales se relacionan los contenidos, una detección automática de reglas esenciales o legales-imperativas de cada documento legal, determinar si el documento legal está correctamente clasificado dejando a un lado sólo una detección estadística de alta exactitud de la clase de documento, una determinación automática de la fecha de emisión del documento, una determinación automática de una lista de nombres de personas naturales o jurídicas encontradas en el texto y otra información relevante mencionada en el texto, que se relaciona con una clase o multiclase de documento, más rápida, completa y precisa que la caracterización manual o la descripción manual realizada por el personal técnico legal, donde esta velocidad y precisión de la comparación es factible cuando se implementa el proceso como un sistema computacional, **CARACTERIZADO** porque el método comprende los pasos de:

a. tras recibir un documento digital, de un usuario conectado a una aplicación web desde su propio computador, procesar dicho documento con una aplicación de Reconocimiento Óptico de Caracteres (OCR, por su sigla en inglés);

b. luego, ejecutar un proceso automático de caracterización de documento dentro de la aplicación y procesar el texto del documento recibido desde el usuario, donde esto implica, en primer lugar, ajustar el texto, por medio del uso de un diccionario de palabras válidas, luego, revisar la aparición de un gran número de diferentes reglas lógicas esenciales con sus coeficientes calibrados, que se asocian a las clases, mediante el reconocimiento de cualquiera de las muchas expresiones diferentes del lenguaje y un conjunto de palabras claves, lo que resulta en una puntuación para cada una de las reglas esenciales, que permiten determinar cuáles clases se aplican a través de un medio de especificación de ciertas reglas lógicas relacionadas con el texto;

c. reconocer los nombres de personas y organizaciones en los contenidos del documento, mediante el uso de una amplia variedad de diferentes expresiones verbales posibles de estos nombres y, de acuerdo con el texto, determinar cuál de estos

nombres son relevantes para los contenidos y cuáles sólo hacen referencias, a través de medios de detección;

d. reconocer y extraer información relevante en función de la definición de clase o multiclase realizada y el conjunto de reglas esenciales definidas para esa clase o multiclase, a través de los medios de detección;

e. reconocer fechas relevantes en el documento y asignar una puntuación a cada una de ellas, de acuerdo con el texto anterior o siguiente, asociado con otras palabras claves y expresiones del lenguaje, que se utilizan para referirse a fechas de emisión de documentos, teniendo en cuenta la clase previamente definida y mediante el uso de las reglas esenciales definidas de la misma clase, a través de los medios de detección, en relación con la clase previamente identificada y el conjunto de reglas que se utilizan para clasificarla; y

f. revisar los contenidos del documento, reconociendo diferentes patrones de texto como una combinación de palabras claves, sinónimos o términos equivalentes, dentro de los contenidos del documento, a través de medios de detección, estableciendo una revisión automática imperativa-lagal de cada documento legal. Utilizando las reglas esenciales.

2. El método de acuerdo con la reivindicación 1, **CARACTERIZADO** porque las reglas se definen por medio del uso de una aplicación de configuración que permite al usuario especificar un conjunto de palabras claves o sinónimos, número de apariciones en los contenidos y/o uso de ciertas frases, definiendo así las reglas.

3. El método de acuerdo con la reivindicación 1, **CARACTERIZADO** porque el proceso de caracterización se calibra por medio del uso de una especificación automática, que define un coeficiente de peso y valor de los elementos en cada regla, esenciales y no esenciales.

4. El método de acuerdo con la reivindicación 1, **CARACTERIZADO** porque las palabras válidas de un lenguaje y su lista de sinónimos a utilizar al ajustar el

texto en un documento dado en formato digital, son el resultado de un proceso de reconocimiento óptico de caracteres.