

MÉTODO DE CARACTERIZACIÓN DE DOCUMENTO

CAMPO TÉCNICO DE LA INVENCIÓN

La presente invención se relaciona con el procesamiento de lenguaje natural, enfocado en la identificación de información relevante en los contenidos de un documento de entrada en formato digital, lo que también se define como caracterización automática de documento, a través de un proceso paso a paso, que utiliza la información previa para buscar y definir los datos que seguirán.

ANTECEDENTES DE LA INVENCIÓN

El estado del arte divulga varios métodos y técnicas relacionados con la clasificación o la caracterización de documentos, según su contenido.

La patente US8380718 (B2) divulga un sistema y método para agrupar documentos similares. Se determinan las frecuencias de apariciones para términos y frases nominales dentro de un conjunto de documentos. Se selecciona un subconjunto de los documentos mediante la eliminación de aquellos documentos que tienen términos y frases nominales, que caen fuera de un rango acotado de condiciones superior e inferior para la frecuencia de aparición.

Por otra parte, la patente europea EP1365329 (B1) divulga un método de clasificación, donde un documento se clasifica en, al menos, una clase de documento, al seleccionar términos para su uso en la clasificación, entre aquellos términos que aparecen en el documento. Una similitud entre el documento de entrada y cada clase se calcula por medio del uso de la información guardada para cada clase de documento. Se corrige la similitud calculada para cada clase. La clase, a la que pertenece el documento de entrada, se determina de acuerdo con la similitud corregida a cada clase.

El documento US2013110843 divulga un método y sistema para clasificar archivos de seguros para la identificación, selección y recolección eficiente de las demandas de subrogación. La invención determina si un reclamo de seguro tiene mérito para justificar los esfuerzos de recuperación del reclamo, que utilizan un código de software para describir parcialmente un conjunto de documentos, que tienen datos de archivos no estructurados y estructurados, que contienen términos y frases con bases contextuales, un código para la transformación de términos y frases, un código para

iterar un proceso de clasificación para determinar las reglas que mejor clasifican el conjunto de documentos, en función del contexto, un código para la incorporación de las reglas en una representación de inducción y conocimiento, taxonomías de tesauros y resumen de textos para clasificar las demandas de subrogación; código para calcular una puntuación base y un vector de concepto para identificar los reclamos seleccionados, que demuestran una determinada probabilidad de recuperación de subrogación.

En primer lugar, los métodos y sistemas descritos anteriormente no permiten un proceso de clasificación múltiple o la caracterización de diferentes elementos de contenido, incluidos en cada documento.

En segundo lugar, todos los métodos descritos anteriormente, a excepción de uno, no consideran formas múltiples dependientes del lenguaje, para la misma expresión semántica (agrupadas en reglón), que implica un modo más flexible para reconocer patrones específicos de texto en el documento, lo que significa que estos métodos dependen de un conjunto de palabras claves, que son sinónimas entre sí. Cada expresión puede estar compuesta por distintas cantidades de palabras, con diferentes sinónimos, conjugaciones, órdenes e, incluso, un conjunto de palabras combinadas, separadas por no más de un número específico de palabras.

En tercer lugar, los métodos descritos anteriormente no miden la importancia de una expresión en un contexto específico, agregando más o menos importancia a aquella expresión para influir en el resultado de la caracterización.

Finalmente, los métodos centrados en la clasificación no abordan un proceso de detección con los contenidos encontrados en los primeros pasos o etapas de la caracterización, conduciendo el comportamiento y el rendimiento de los siguientes pasos de la caracterización.

RESUMEN DE LA INVENCION

La presente invención se relaciona con un procesamiento de lenguaje natural, enfocado en la identificación de elementos de información específica en los contenidos de un documento de entrada en formato digital, lo que también se define como caracterización automática de documento, a través de un proceso paso a paso, que utiliza la información previa para buscar y definir los datos que seguirán. Este

enfoque de procesamiento de texto se centra en documentos, en los que diferentes elementos de información son relevantes para identificarse, a partir de contenidos que no siguen un diseño, orden o sintaxis específico, conocidos como documentos no estructurados. Los documentos no estructurados, como documentos legales, no presentan una secuencia específica de elementos y los contenidos se pueden encontrar utilizando diferentes expresiones, según el lenguaje utilizado.

Los elementos de información que se buscan en el documento permiten determinar las características específicas de este documento. Este es el caso de las clases o categorías de documentos y otra característica asociada, tales como fechas, nombres de personas naturales o jurídicas mencionadas en el texto y otra información pertinente mencionada en el texto, que se relaciona con la clase o multiclase de documento.

Aunque la forma en que se buscan estos elementos se diferencia de otras, en todas ellas el enfoque es reconocer patrones de texto específicos dentro de los contenidos del documento y conducir el proceso de detección con los contenidos encontrados en los primeros pasos de la caracterización, utilizando el conjunto de reglas definido para cada clase.

El método para la caracterización automática de documentos que aquí se describe incluye la clasificación de documentos, que es un proceso secuencial para asociar documentos a clases pre-definidas de documento, y ha sido abordada por diferentes métodos y técnicas, como la Naïve Bayes y las máquinas de vectores de soporte, entre otros. Sin embargo, estos métodos no se preparan adecuadamente para la acumulación natural de las clases, la que, en este tipo de documentos, no sólo es factible, sino que esperable. Además, estos métodos, a excepción de uno, se centran en las palabras claves específicas, no considerando fácilmente diferentes expresiones que dependen del idioma, como conceptos o ideas equivalentes.

El proceso para la caracterización automática de documentos procesa cinco veces los contenidos de documento (reconocimiento óptico de caracteres (OCR, por su sigla en inglés), determinación de clase y multiclase, detección de nombres, caracterización de documento relacionada con la clase, detección de fecha de documento), ya que se necesitan enfoques diferentes para cada uno de los elementos de información.

La primera revisión tiene como objetivo refinar el texto de entrada, reconociendo que los resultados del Reconocimiento Óptico de Caracteres (OCR, por su sigla en inglés) pueden separar palabras completas, al asumir espacios y final de las líneas. Un diccionario de palabras válidas en un lenguaje determinado se utiliza para validar las combinaciones de palabras; de esa forma, se corrigen palabras separadas, permitiendo que el resto de las revisiones se ejecuten con facilidad.

La segunda revisión se centra en las clases de documentos, mediante la búsqueda de reglas semánticas que pueden aparecer en el texto y, de acuerdo con las reglas de aparición, determinar una o más clases que puedan aplicarse.

La tercera revisión se centra en los nombres a los que se hace referencia en los contenidos. Esta revisión utiliza una lista de palabras claves relevantes, que podrían incluirse tanto al principio como al final de un nombre, y también que lo podría preceder o seguir, pero no se encuentran incluidas en el nombre. La relevancia de cada una de las palabras claves definidas puede estar asociada a la clase principal asignada al documento en el paso anterior; por lo tanto, permite un proceso de detección bien afinado.

La cuarta revisión está relacionada con la caracterización detallada del documento, basado en la clase o multiclase que ha sido previamente definida. Dada esta clase o multiclase, existe una lista de las características que tiene un documento. A modo de ejemplo, en el caso de un contrato de arrendamiento inmobiliario, es útil conocer la "renta de arrendamiento" o la "superficie alquilable" por lo que se debe realizar la búsqueda de esta información.

La revisión final busca sólo fechas válidas, pero también considera algo del texto precedente y la información definida, para ayudar en la determinación de la fecha de emisión. Hay que tener en cuenta que cada clase puede tener una o más fechas relevantes, por ejemplo, tal como la fecha en la que comienza un contrato de arrendamiento y la fecha en la que culmina. En este paso, se pre-establecen las fechas a buscar, ya que se conoce la clase de documento.

Esta invención tiene como objetivo producir un resultado de caracterización más completo, preciso y rápido, en comparación con la caracterización manual que un experto legal puede proporcionar al leer los contenidos del documento. Esta comparación de velocidad y precisión es posible cuando se implementa el proceso

como un sistema computacional, disponible para la caracterización de documentos para los usuarios de este sistema. La precisión en la caracterización es apoyada por un proceso de calibración de regla, que considera un conjunto de documentos previamente clasificados, y ajusta los coeficientes de cada regla para que coincidan con estas clases de documentos de la mejor manera posible, lo que permite la mejora continua de esta exactitud.

BREVE DESCRIPCIÓN DE LAS FIGURAS

La FIG. 1 es un diagrama en bloques, que representa el proceso de caracterización de documento.

La FIG. 2a es un diagrama en bloques, que representa el proceso de determinación de clase y multiclases.

La FIG. 2b es un diagrama en bloques, que representa el proceso de determinación de clase y multiclases, con reglas esenciales.

La FIG. 3 es un diagrama en bloques, que representa un ejemplo de una modalidad de la invención. En este escenario en particular, la regla de “compra existente” se utilizó como un ejemplo.

La FIG. 4 es un diagrama en bloques, que representa el proceso de calibración de clase y multiclases.

La FIG. 5 es un diagrama en bloques, que representa el proceso de detección de nombre.

La FIG. 6 es un diagrama en bloques, que representa la caracterización relacionada con clase de documento.

La FIG. 7 es un diagrama en bloques, que representa el proceso de detección de datos de un documento.

La FIG. 8 es un diagrama en bloques, que describe un sistema computacional para ejecutar la caracterización de documento, desde la recepción del documento a caracterizar, la recuperación de palabras desde el diccionario y de reglas textuales desde una base de datos, hasta la visualización final de las características resultantes del documento proporcionado.

DESCRIPCIÓN DETALLADA DE LAS MODALIDADES DE LA INVENCIÓN

En esta sección se detalla y explica cada uno de los procesos, de tal manera que la comprensión del proceso se realice de una forma más sencilla. Cada proceso tiene su propia figura (FIG. **2a**, **2b**, **3**, **4**, **5**, **6**, **7**) con cada explicación y se incluye también una figura del proceso completo (FIG. **1**).

Proceso principal

El siguiente texto utiliza la FIG. 1 para explicar todo el proceso. Cada proceso tiene su propia sección, destinada a la explicación de cada proceso, a continuación de este párrafo. Como se mencionó anteriormente, el primer paso es la revisión del texto con el reconocimiento de palabras individuales **101** con un Reconocimiento Óptico de Caracteres (OCR). Una vez que el texto ha sido revisado por el OCR, se realiza la determinación de clase o multiclase **102**, alineada con un proceso de calibración de regla de clase o multiclase **102a**, que es un proceso iterativo, que mejora la definición de puntuación de regla, obteniéndose así una mejor clasificación. Con la clase ya definida, se completa un proceso de detección de nombres **103**, con el objetivo de obtener nombres relevantes de personas y organizaciones en el documento. Este proceso **103** considera la clase a la que el documento se ha asociado para hacer la búsqueda **105** de una manera más precisa, teniendo en cuenta un conjunto relevante de palabras claves y expresiones **103a** para definir los nombres contenidos en el documento. Una vez que se han definido las clases y los nombres, es posible iniciar una caracterización de documento relacionada con la clase **104**, que busca información teniendo en cuenta la clase definida previamente. El paso final consiste en la búsqueda y definición de la fecha de emisión válida del documento y otras fechas relevantes **106**, que de modo similar al proceso de la búsqueda de nombres, toma en consideración la clase que se ha definida para realizar la búsqueda. Una vez que han terminado todos estos procesos, se ha llevado a cabo la caracterización final de documento **107**. Esta caracterización final de documento **107** incluye fechas, nombres de personas naturales o jurídicas mencionadas en el texto y otra información pertinente mencionada en el texto, por lo que se logra una descripción detallada del documento, de una manera rápida y precisa. Esta comparación de velocidad y precisión es posible cuando se

implementa el proceso como un sistema computacional, disponible para la caracterización de documentos para los usuarios de este sistema.

Para más detalles de cada uno de los procesos, favor remitirse a la sección correspondiente.

Ajuste de texto de documento

El texto del documento se revisa a través del reconocimiento de palabras individuales, lo que se obtiene mediante un medio de procesamiento de Reconocimiento Óptico de Caracteres (OCR) **101**. Si el origen del documento es un procesador de palabras, el proceso de ajuste puede producir una diferencia mínima e, incluso, no producir diferencia alguna. Para intentar corregir la detección de espacios dentro de las palabras a través del proceso de OCR, cada palabra individual se concatena con la siguiente y después se compara con la lista de palabras válidas en el lenguaje del documento almacenado en el diccionario **101a**. Si se encuentra esta palabra concatenada, entonces ésta se mantiene durante los siguientes procesos.

Determinación de clase o multiclase

El proceso depende de una pluralidad de reglas definidas por medio de la especificación de reglas lógicas relacionadas con el texto **301**, donde cada uno de ellos define un conjunto de palabras claves que representan las expresiones verbales equivalentes **301a**. Estas expresiones se describen a través de palabras claves, sinónimos, conjugaciones, diferente cantidad de palabras y/o expresiones verbales equivalentes **301a** del mismo concepto, también considerando los diferentes niveles de separación entre estas palabras, permitiendo así una mayor flexibilidad en la detección de regla de texto (**2a.201** y **2b .201**). Cada regla se puede relacionar a una o más clases (**2a.202** y **2b.202a**), pero la relevancia de cada clase está determinada por un coeficiente de relevancia, que puede tener un valor positivo o negativo **305**. En algunas clases, una regla puede ser considerada como esencial, en cuyo caso el coeficiente debe ser alto e, incluso, la ausencia de esta regla podría dar lugar a una penalización con respecto a la clase de documento, en el que esta regla se considera esencial **2b.202b**. Esto significa que la misma regla podría ser muy relevante para una clase de documentos, al utilizar un valor de coeficiente positivo bastante alto, incluso siendo

considerada como esencial; también podría ser ligeramente relevante para una segunda clase, que puede ser referenciada por un menor valor de coeficiente positivo; también podría no tener relevancia alguna, lo que se representa por un coeficiente nulo; incluso, podría ser considerada como contradictoria para una clase de documentos, lo que se representa a través de un valor de coeficiente negativo **2b.202b**.

En la FIG. 2b se ilustra la dualidad de una regla esencial **2b.202b**, que se debe a diferentes valores que puede tener para distintas clases. Para una clase dada, tal como un documento de compra inmobiliaria, la regla de la “compra existente” puede ser relevante para indicar que el documento es efectivamente un documento de compra. Por otra parte, un documento, tal como un acuerdo de arrendamiento inmobiliario, no tendrá esta regla como esencial y la ponderación que esta regla tiene en la puntuación total del documento puede ser muy baja (**2a.203** y **2b.203**).

Por ejemplo, en la modalidad preferida de este método, la caracterización de documentos legales, la regla de la “compra existente” **301** se aplica cualquier tipo de compra, tal como de bienes raíces, muebles o acciones. En cualquiera de estos diferentes tipos de acuerdos, esta regla puede encontrarse en formas ligeramente diferentes **301a**. Una posible expresión que se relaciona con esta regla es "... las partes acuerdan suscribir el presente acuerdo de compra ..." **301b**. Una expresión alternativa es "... el vendedor vende... y el comprador compra..." **301b**, independientemente del ítem considerado en la venta. La metodología para todas las demás reglas funciona de forma similar, implementando un conjunto de palabras claves que representan términos equivalentes (**302** y **304**).

Una regla adicional y complementaria que se aplica a este tipo de contratos es el "precio de compra existente", que se refiere al precio que el comprador va a pagar. Con respecto a la regla del "precio de compra existente", ésta se puede encontrar en diferentes formas en el texto, tal como "... el precio pagado por el comprador es el siguiente..." o "... el precio de... será de...".

En un ejemplo relacionado, un tipo diferente de contrato es el acuerdo o contrato de arrendamiento. La regla principal, que se encuentra en este tipo de acuerdos o contratos, es el "arrendamiento existente" que se puede encontrar como "... el arrendador arrienda por la presente las instalaciones al arrendatario...", y más tarde en el documento "... el arrendatario arrienda las instalaciones", que corresponden a un

acuerdo o contrato de arrendamiento de bien raíz. Por exactamente esta misma regla, una forma diferente de expresarlo es "... el arrendatario recibe del arrendador...". En el mismo tipo de documento, otra regla que se puede encontrar es la "duración del arrendamiento", que describe la duración del contrato. Esta regla se puede encontrar en el documento como "... este arrendamiento se deberá iniciar el [fecha1] y finalizará el [fecha2]", o también como "... este arrendamiento estará vigente desde el [fecha] con una duración de X días...". Este conjunto de reglas se encuentran en los contratos de arrendamiento, por lo tanto, se deben ajustar coeficientes bastante altos para asociarlos a la clase de contrato de arrendamiento, e incluso considerando uno de ellos o ambos como esenciales, asignándoles así el coeficiente más alto para esta clase (**303**).

El proceso de clasificación busca el documento desde el principio hasta el final, identificando cada palabra individualmente y comparándola con una de las varias reglas **301a** definidas. Si la palabra coincide con alguna de las palabras de cualquiera de las reglas, entonces, el siguiente texto se procesa para completar cada una de esas reglas candidatas que están siendo evaluadas, sino que también compara cada nueva palabra con otras reglas **301b**.

El proceso de clasificación termina con un conjunto de reglas encontradas para aplicar a cada documento (**2a.202** y **2b.202**). Entonces, la puntuación de clasificación final para cada clase de documento se calcula según la siguiente fórmula **305**.

$$\text{ClasePuntuación}(\text{documento}, \text{regla}) = \text{apariciones} * \text{Coeficiente}$$

Esto significa que si el coeficiente de cierta regla es alto para una clase de documento, alguna de las apariciones repetidas se sumará para asignar esa clase al documento.

La definición de las puntuaciones anteriores para cada clase tendrá efecto sobre la calibración de clase o multiclase (**306**, **2a.204** y **2b.204**), y luego este proceso iterativo (**2a.205** y **2b.205**) va a modificar y mejorar el conjunto de reglas que se utilizan para clasificar el documento.

Detección de nombres

Cualquier documento, especialmente documentos legales, puede hacer referencia a nombres de personas naturales y organizaciones. Por ello, este proceso se centra en la producción de una lista de todos los nombres encontrados en los contenidos **501**. Esta detección, por medio del uso de medios de detección, se basa en el hecho de que los documentos mencionan nombres de personas, precedidos por algunas expresiones verbales específicas **502a**, tales como "Señor" o "Señora" y el nombre podría ser seguido por una referencia de identificación personal o un verbo, que especifica una acción con referencia a esta persona. En el caso de organizaciones, el texto precedente también podría incluir expresiones equivalentes y, por lo general, termina con "SRL", "S.A." u otras siglas tipo de compañías, que reglalmente se encuentran en los documentos legales.

Por lo tanto, estos pasos consideran una lista de expresiones precedentes, que determinan que el texto que sigue es un nombre, pero no están incluidos en el propio nombre, tal como un "señor" o "señora" y también una lista de las subsiguientes expresiones, que podrían también estar incluidas o no en el nombre, y que se utilizan para encontrar el final del nombre de referencia. Este paso y la lista de las expresiones anteriores y posteriores se incluyen en una aplicación **502**.

Para mejorar el proceso de detección de nombres, se pueden considerar algunas reglas adicionales, en función de las clases de documentos determinadas en los pasos anteriores **501a**. Por ejemplo, un documento asociado con la clase de documento "contrato de compra" que está determinado por la aparición de la regla "compra existente" y otras, puede incluir expresiones tales como "... como el vendedor ..." o "... también conocido como el comprador ... ", como texto que sigue a los nombres en el documento. Estas reglas adicionales que dependen de la clase de documentos mejoran la detección del nombre completo, especialmente en la detección del final de un nombre. También es posible que, teniendo en cuenta la clase definida previamente, haya un reconocimiento de otros nombres relevantes o no relevantes en el documento. Este tipo de clase previamente definida tendrá un gran impacto en la forma en que se realiza el mencionado proceso **502b**.

Caracterización de documento relacionada con la clase

Esta caracterización se refiere a la detección de otros datos específicos del documento, tales como cantidades numéricas, nombres propios, entre otros. En algunos casos, es de gran utilidad obtener información particular de un documento, tal como el alquiler de arrendamiento en un contrato de arrendamiento de bienes raíces o la superficie que se puede arrendar. Una vez que la clase o multiclase ya ha sido definida **602a** y la búsqueda de nombres ha terminado **601**, se define la lista de características de un documento en particular. Dada esa lista, se busca la información del documento **602**. Esto se hace de gran utilidad para la caracterización **603** del documento, ya que toma en consideración la clase del documento, para elaborar una caracterización completa del mismo. Un ejemplo de esta definición es el acuerdo o contrato de arrendamiento de bien raíz **603a**. En este documento, es útil conocer el precio o alquiler del arrendamiento **603b**, por lo que la búsqueda de esta información se completa una vez que la clase ha sido previamente definida. El uso de las reglas definidas, dada la clase o multiclases, resulta muy útil, ya que la aparición de reglas en el texto es una clara evidencia de que la información relacionada con la clase debe aparecer, por lo que una vez que la regla ha sido identificada en el texto, es posible extraer información relacionada con la regla. Es relevante mencionar que este uso en particular se puede replicar en otros documentos y su información particular. Este proceso se explica en la FIG. 6.

Detección de datos del documento

Por medio del uso de medios de detección, el paso final es la detección de la fecha de emisión del documento, considerando que se puede incluir más de una fecha válida en el documento, pero sólo una de ellas puede ser considerada como la fecha de emisión del documento. Además, esta detección considera la existencia de muchas formas verbales para describir una fecha **701**.

Tomando en cuenta estas consideraciones, el proceso de detección de la fecha del documento se realiza mediante el reconocimiento de cada fecha válida y el cálculo de una puntuación a través de los medios de detección **702**, que depende de los siguientes criterios: la fecha de emisión del documento debe ser anterior a la fecha

actual, debe estar completa, lo que significa que debe incluir el día, mes y año, y puede estar precedida por una expresión que hace referencia a una ubicación.

La fecha de emisión definitiva del documento es la que obtiene la puntuación más alta y la primera, en caso de empate **703**.

Este proceso, las técnicas para detectar fechas válidas en el documento, y la lista de palabras claves que permiten la determinación de la fecha de emisión del documento se pueden configurar en una aplicación.

Dado que la clase o multiclase del documento ha sido definida **704**, es necesario buscar una o más fechas que sean relevantes en el documento. Las reglas usadas anteriormente para definir el documento se utilizan en la búsqueda de las fechas claves, ya que, en algunos casos, el texto correspondiente a la regla tiene la información pertinente en las cercanías. Las fechas relevantes del documento ya serán definidas con una lista preestablecida. Luego, la búsqueda de estas fechas se hará **705**, completando el proceso de caracterización del documento **706**. A modo de ejemplo, en un contrato de arrendamiento de bienes raíces, es útil conocer la fecha en la que comienza el arrendamiento, así como la fecha en que el contrato de arrendamiento termina.

Calibración de clase o multiclase

Para mejorar la calidad del método de determinación de clase o multiclase, existe un proceso de formación que utiliza medios de especificación automática, que recibe un conjunto de documentos previamente clasificados **401**, definido como el conjunto de formación, y luego procede a asignar diferentes coeficientes a todas las reglas definidas, con el objetivo de igualar la aparición de ellas en los documentos del conjunto de formación **403**, y también intentar coincidir -lo más cercano posible- con las clases asignadas a cada una de ellas **402**. Este proceso utiliza un Algoritmo Genético (**404**), también conocido como AG, con una función objetivo que define la idoneidad de una combinación específica de los coeficientes asignados a las reglas, que se basa en la medición de las diferencias entre las clases pre-asignadas a los documentos del conjunto de formación **401**, y las clases asignadas de forma automática con una combinación candidata determinada de coeficientes para cada regla y cada documento. Luego, reducir esta diferencia de forma iterativa para llegar a

un conjunto final mejorado de coeficientes **405**. La mejora del conjunto de los coeficientes es un proceso constante e iterativo, que permitirá caracterizar los documentos legales de una manera correcta.

La Figura 8 ilustra una modalidad ejemplar de un sistema computacional basado en un servidor web **800**, que almacena y ejecuta una modalidad de la presente invención. Este servidor web almacena una aplicación para implementar el método de caracterización y una aplicación para el OCR **803**, que procesará los documentos recibidos para su caracterización. La aplicación web **801** es la principal aplicación a la que los usuarios tienen acceso para ejecutar tanto los procesos de caracterización como los de OCR, al conectarse desde un computador de usuario **808**, por medio del uso de una red **807**, y el uso de esta aplicación web para caracterizar los documentos **809**, que los usuarios tienen en su propio local disco del computador.

El servidor web reacciona a una petición del usuario para la caracterización de documentos, al recibir el documento, procesarlo mediante OCR y recuperar el diccionario de palabras válidas **802**, almacenadas en un archivo en el disco del servidor local, a la memoria local **804**, y también recuperar las reglas de recolección y las clases predefinidas de documentos, almacenados en una base de datos **806** accesible para el servidor. El proceso de caracterización se ejecuta entonces por la CPU **805**, lo que resulta en una página web con la información obtenida por la caracterización del documento suministrado, que se envía de nuevo al computador del usuario y se muestra en su pantalla o monitor.

El proceso de calibración opera de modo similar. Un usuario administrador carga en el servidor web varios documentos previamente caracterizados y ejecuta el proceso de calibración mediante otra opción, en la misma aplicación web. Este proceso de calibración, finalmente, modifica los coeficientes para las reglas, actualizándolas en la base de datos.