

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 371 918**

51 Int. Cl.:  
**G06F 17/24** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **05103224 .1**  
96 Fecha de presentación: **21.04.2005**  
97 Número de publicación de la solicitud: **1596310**  
97 Fecha de publicación de la solicitud: **16.11.2005**

54 Título: **CUMPLIMENTACIÓN AUTOMÁTICA E INTELIGENTE DE FORMULARIOS.**

30 Prioridad:  
**12.05.2004 US 844540**

45 Fecha de publicación de la mención BOPI:  
**11.01.2012**

45 Fecha de la publicación del folleto de la patente:  
**11.01.2012**

73 Titular/es:  
**MICROSOFT CORPORATION  
ONE MICROSOFT WAY  
REDMOND, WA 98052, US**

72 Inventor/es:  
**Kadie, Carl M.;  
Goodman, Joshua T.;  
Bradford, Donald E.;  
Chickering, David M. y  
Glasgow, Dane A.**

74 Agente: **Carpintero López, Mario**

**ES 2 371 918 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

Cumplimentación automática e inteligente de formularios

La presente invención versa en general acerca del aprendizaje de información de usuario y, en particular, al establecimiento de una correlación entre la entrada por parte del usuario y uno o más campos de un formulario basado en la red para facilitar la entrada automática de la información del usuario.

La comunicación y la realización de negocios por Internet se ha convertido en una experiencia cotidiana tanto para profesionales en el ejercicio de su cometido como para el público consumidor. A los usuarios de Internet se les pide comúnmente que cumplimenten numerosos formularios —algunos de los cuales pueden ser similares o no—, como cuando se compran productos, se solicita información o se formulan preguntas. Los usuarios deben introducir de forma reiterada la misma información, incluyendo su nombre, su dirección particular, su dirección comercial, su dirección de correo electrónico y/o números de teléfono en una variedad de formularios diferentes.

Desgraciadamente, los procedimientos convencionales que han intentado abordar estas inquietudes son problemáticos. Por ejemplo, en algunas técnicas, pueden recordarse las entradas previas, pero solo si el formulario es idéntico al formulario previo. Considerando el amplio abanico de sitios web y el enorme número de propietarios y operadores de sitios web en Internet, esta puede no ser una técnica o una solución razonables para minimizar la pérdida de tiempo del usuario y aumentar la eficiencia del usuario en Internet.

Otros procedimientos tradicionales pueden proporcionar una introducción automatizada de algunos datos; sin embargo, estos procedimientos pueden seguir siendo deficientes y, por lo tanto, incapaces de satisfacer las muchas necesidades del usuario de Internet de hoy.

CHUSHO T ET AL: "Automatic filling in a form by an agent for web applications", SOFTWARE ENGINEERING CONFERENCE, NINTH ASIA-PACIFIC, 4-6 de diciembre de 2002, PISCATA-WAY, Nueva Jersey, IEEE, 4 de diciembre de 2002 (2002-12-04), páginas 239-247, describen una tecnología para cumplimentar automáticamente un formulario por medio de un agente de usuario en colaboración con un agente. En particular, se usan reglas cognitivas basadas en el diseño del formulario, se introduce una ontología de nombres conceptuales para diferentes expresiones del mismo significado y se emplean reglas experimentales de otros usuarios para mejorar la tasa de éxito.

Es el objeto de la presente invención proporcionar un sistema mejorado de cumplimentación automática.

El objeto se resuelve por la materia de las reivindicaciones independientes.

Las realizaciones preferentes de la presente invención están definidas por las reivindicaciones dependientes.

Lo que sigue presenta un resumen simplificado de la invención para proporcionar una comprensión básica de algunos aspectos de la invención. Este resumen no es una visión general amplia de la invención. No se pretende que identifique elementos clave/críticos de la invención ni que delinee el alcance de la invención. Su único propósito es presentar algunos conceptos de la invención de forma simplificada como preludeo de la descripción más detallada que se presenta después.

La presente invención versa acerca de una técnica mejorada de entrada de datos para su uso por parte de un navegador en una variedad de formularios web diferentes que minimiza el esfuerzo del usuario. Más específicamente, la invención permite un sistema y/o una metodología que puede emplear técnicas de aprendizaje de máquina que cumplimentar automáticamente ( ) uno o más campos en un conjunto diversos de formularios web. Usando técnicas de aprendizaje de máquina, el navegador u otra herramienta pueden aprender qué datos o tipos de datos se correlacionan con qué campos o tipos de campos. También pueden aprenderse otros tipos de correlaciones, como que datos particulares se introducen en campos particulares en una página web particular. Por ejemplo, la o las técnicas de aprendizaje de la presente invención pueden aumentar la cumplimentación automática de datos en comparación con la programación explícita.

Según un enfoque, el sistema y el procedimiento pueden utilizar una o más bases de datos de información sobre el usuario y aprender qué campos se correlacionan con qué entradas (o elementos) de la base de datos. En particular, puede generarse una base de datos de información de campos, en parte, observando y reuniendo lo introducido por el usuario por medio de una herramienta o barra de herramientas . También puede aprenderse a determinar qué campos de formulario (campos de un formulario basado en la red) debería cumplimentarse con qué campos de la base de datos (campos procedentes de la base de datos). Por ejemplo, puede aprender que en un campo de formulario etiquetado NOMBRE es probable que el usuario introduzca la información del "nombre de pila" de su base de datos. Cuando un usuario llega a un nuevo formulario web, el navegador, por ejemplo, precisa determinar si el formulario incluye entrada que puedan ser cumplimentadas por el navegador y, en caso afirmativo, qué entradas corresponden a qué partes de la base de datos. Además, si hay múltiples entradas en la base de datos, entonces también debería efectuarse la determinación de qué entrada de la base de datos usar para una formulario dado.

Obtener los datos de entrenamiento para realizar cualquiera de las determinaciones anteriores puede ser una tarea crucial. Afortunadamente, no es preciso que el entrenamiento sea necesariamente específico del usuario. En un aspecto de la invención, los datos de entrenamiento que comprenden las características correctas de entrada y salida pueden ser reunidos de alguna manera. Por ejemplo, una barra de herramientas puede reunir y agregar para algunos usuarios qué entradas de la base de datos pone en qué campos y cuáles son las características de esos campos. Esta información puede ser devuelta a un repositorio central. Los datos del repositorio central pueden ser usados como entrada a un algoritmo de entrenamiento para aprender qué campos de la base de datos corresponden a los campos del formulario. Por ejemplo, puede aprenderse que un campo de la base de datos etiquetado “nombredepila” se corresponde con un campo del formulario etiquetado NOMBRE. Si un usuario introduce “Joshua” en el campo del formulario etiquetado NOMBRE y tiene un campo de la base de datos etiquetado “nombredepila”, entonces puede devolverse la información NOMBRE y “nombredepila”. Así, no tiene por qué incluirse información personalmente identificable para llevar a cabo este aspecto de la invención.

Tal como se ha mencionado más arriba, pueden emplearse los campos de la base de datos y diversas características derivadas de los mismos con fines de entrenamiento. Por ejemplo, pueden explotarse las características de los campos de formularios y ser usadas como características para distinguir entre diferentes tipos de datos y diferentes tipos de campos de formulario. En particular, cada campo de formulario tiene un tipo de entrada, como texto, que puede indicar un cuadro de texto. El hecho de que el campo del formulario sea un campo de tipo texto puede ser una información útil, porque puede facilitar deducir qué datos están incluidos en el campo. Es decir, si fuese algo distinto de campo de texto, por ejemplo un cuadro combinado, probablemente no contendría un nombre ni una dirección.

Las características puede estar basadas en los nombres de los campos del formulario, en el tamaño en caracteres o la anchura del campo del formulario, en el texto que rodea al campo del formulario y/o en la posición del campo del formulario con respecto a otros campos del formulario que pueden o no ser conocidos o identificados. El examen de las tablas HTML también puede facilitar determinar qué palabras o datos se sitúan encima, debajo, a la izquierda o a la derecha de un campo particular del formulario. Además, usar texto cercano o subcadenas de texto cercano que pueden o no ser idénticas al texto cercano puede proporcionar información adicional.

En otro aspecto de la invención, pueden asignarse campos de la base de datos a campos o valores del formulario en base, al menos en parte, a una o más restricciones. Las restricciones pueden controlar las asignaciones de los campos o las entradas de la base de datos para formar valores en muchas circunstancias diferentes. Por ejemplo, pueden emplearse restricciones cuando es evidente que el formulario web tiene tipos de salida contradictorios o para verificar que las entradas de la base de datos pueden o no ser reutilizadas dentro de un único formulario web.

En algunos casos, como cuando se nuevo formulario web, puede desearse una asignación óptima de entradas de la base de datos a los campos del formulario. La mejor asignación puede determinarse, en parte, maximizando las probabilidades conjuntas según una o más restricciones. Ejemplos de los tipos de restricciones pueden incluir no permitir un campo del formulario (por ejemplo, la 2ª línea de la dirección) sin otro campo del formulario (por ejemplo, la 1ª línea de la dirección), reutilizar las restricciones para mitigar el uso reiterado de una entrada, restricciones de composición para gestionar tipos de salida contradictorios y/o restricciones de orden (por ejemplo, el prefijo local precede a la central telefónica).

En otro aspecto adicional de la presente invención, puede personalizarse la correspondencia entre los campos de la base de datos y los valores del formulario. Por ejemplo, puede observarse que en una página web particular, un usuario rechaza la entrada automática sugerida e introduce datos manualmente. Parte de estos datos puede coincidir con entradas en la base de datos. A partir de entonces, puede aprenderse que, para este usuario particular, ciertas entradas de campos del formulario corresponden a las entradas observadas de la base de datos, no a las entradas aprendidas de la base de datos. De ahí que el usuario pueda invalidar la entrada automática con algunos otros datos y el sistema o el procedimiento puedan aprender estos otros datos. De forma alternativa o adicional, las entradas manuales del usuario pueden ser específicas al sitio (para este usuario) —lo que significa que tal entrada se usa únicamente para esta página web o este sitio web particular—. Además, las entradas automáticas pueden ser modificadas en base a la entrada observada del usuario en respuesta a una operación de cumplimentación automática.

En otro aspecto adicional de la invención, pueden emplearse la heurística o el aprendizaje cuando hay múltiples conjuntos de campos, como dirección particular y dirección comercial o dirección de facturación y facturación de envío. Por ejemplo, en los formularios HTML, estas direcciones aparecen a menudo en tablas diferentes. Así, sus respectivas ubicaciones o tablas pueden ser aprendidas y pueden ser separadas en cualquier número de subconjuntos.

Otro enfoque puede no usar una base de datos, sino que, en su lugar, implica observar y/o hacer seguimiento de la entrada de datos del usuario para aprender qué datos introducir en qué campos. En particular, la entrada de datos del usuario en una pluralidad de campos del formulario puede ser observada y almacenada para un aprendizaje desconectado de la línea o ser objeto de seguimiento y adaptada para el aprendizaje con conexión a la línea. En el aprendizaje sin conexión a la línea, puede almacenarse una pluralidad de casos observados de entradas de

campos, incluyendo información sensible, y usarse luego como entrada para el sistema de aprendizaje de máquina en un momento posterior. El aprendizaje con conexión a la línea emplean un algoritmo de aprendizaje con conexión a la línea, como uno de los algoritmos de enfoque bayesiano ingenuo, de descenso de gradiente, Winnow y preceptrón, entre otros, que no necesite almacenar ninguna información, salvo, por ejemplo, el propio modelo que puede contener alguna información personalmente identificable. Muchos aspectos de la invención expuestos en lo que antecede también pueden ser combinados con este enfoque de las entradas observadas.

En algunos casos, puede que no todos los campos del formulario sean visibles para el usuario; no obstante, pueden ser cumplimentados. Esto puede suscitar muchas inquietudes de seguridad y privacidad por el mero hecho de que puede no estar al tanto de la información que está publicando. En respuesta a tales inquietudes, otro aspecto de la presente invención proporciona una visualización en la interfaz de usuario que puede permitir que un usuario vea una enumeración de los campos del formulario en la página y/o de los que fueron cumplimentados de forma automática. Esto puede resultar especialmente efectivo para los botones de radio o las casillas de control cumplimentados automáticamente, dado que pueden no ser perceptibles al usuario o el usuario puede no esperar que se cumplimenten automáticamente.

La enumeración de los campos del formulario también puede estar vinculada a los campos reales del formulario en la página web. De ahí que una modificación a un campo enumerado del formulario pueda afectar al contenido en el campo real del formulario. De forma alternativa, el uso de la característica de cumplimentación automática puede estar ligada a medidas de seguridad implementadas en el ordenador del usuario. Es decir, la característica de cumplimentación automática puede estar disponible únicamente en sitios de confianza o en sitios que satisfagan el nivel de seguridad del usuario.

Para el logro de los fines precedentes y relacionados, en el presente documento se describen ciertos aspectos ilustrativos de la invención en conexión con la siguiente descripción y los dibujos adjuntos. Sin embargo, estos aspectos son indicativos únicamente de algunas de las diversas maneras en las que pueden emplearse los principios de la invención, y se pretende que la presente invención incluya todos los aspectos de ese tipo y sus equivalentes. Otras ventajas y características novedosas de la invención pueden hacerse evidentes a partir de la siguientes descripción detallada de la invención cuando se la considera en conjunto con los dibujos.

**Breve descripción de los dibujos**

La Fig. 1 es un diagrama de bloques de alto nivel de un sistema inteligente de cumplimentación automática que hace uso del aprendizaje de máquina según un aspecto de la presente invención.

La Fig. 2 es un diagrama de bloques de un sistema inteligente de cumplimentación automática que suministra datos recogidos a una o varias bases de datos, en las que pueden ser usados como entrada para un sistema de aprendizaje de máquina para que aprenda qué datos corresponden a qué campos según otro aspecto de la presente invención.

La Fig. 3 es un diagrama de bloques de un sistema inteligente de cumplimentación automática que aprende qué datos corresponden a qué campos, en parte, observando las entradas del usuario según otro aspecto adicional de la presente invención.

La Fig. 4 es una visualización ejemplar de pantalla de un formulario web al que puede aplicarse un sistema inteligente de cumplimentación automática para cumplimentar efectiva y eficientemente los campos deseados en base, en parte, a los datos aprendidos según un aspecto de la presente invención.

La Fig. 5 es un diagrama esquemático de bloques que demuestra el empleo de un sistema inteligente de cumplimentación automática con respecto a un nuevo formulario web según otro aspecto adicional de la presente invención.

La Fig. 6 es un diagrama de flujo que ilustra una metodología ejemplar que facilita la entrada automática de campos de un formulario en base a técnicas de aprendizaje de máquina según un aspecto de la presente invención.

La Fig. 7 es un diagrama de flujo que ilustra una metodología ejemplar que facilita aprender qué datos corresponden a qué campos según un aspecto de la presente invención.

La Fig. 8 es un diagrama de flujo que puede extenderse de la metodología de la Fig. 7 que ilustra un enfoque ejemplar al aprendizaje de la entrada del usuario según un aspecto de la presente invención.

La Fig. 9 es un diagrama de flujo de una metodología ejemplar que facilita la introducción de datos de usuario en un nuevo formulario web en base, en parte, a cualquier número de restricciones y condiciones según un aspecto de la presente invención.

La Fig. 10 es un diagrama de flujo de una metodología ejemplar que facilita la conservación de la entrada del usuario introducida encima de la entrada cumplimentada automáticamente según un aspecto de la presente invención.

La Fig. 11 ilustra un entorno ejemplar para la implementación de diversos aspectos de la invención.

5 **Descripción detallada de la invención**

La presente invención se describe ahora con referencia a los dibujos, en los que se usan números de referencia semejantes para referirse a elementos semejantes de principio a fin. En la siguiente descripción, con fines de explicación, se exponen numerosos detalles específicos para proporcionar una comprensión cabal de la presente invención. Sin embargo, puede resultar evidente que la presente invención puede ser puesta en práctica sin estos detalles específicos. En otros casos, se muestran en diagrama de bloque estructuras y dispositivos bien conocidos para facilitar la descripción de la presente invención.

Según se usan en esta solicitud, se pretende que los términos “componente” y “sistema” se refieran a una entidad relacionada con los ordenadores, ya sea un soporte físico, una combinación de soporte físico y soporte lógico, soporte lógico o soporte lógico en ejecución. Por ejemplo, un componente puede ser, pero no está limitado a ser, un proceso que se está ejecutando en un procesador, un procesador, un objeto, un ejecutable, un hilo de ejecución, un programa y un ordenador. A título de ilustración, tanto una aplicación que se está ejecutando en un servidor como el servidor pueden ser un componente. Uno o más componentes pueden residir dentro de un proceso y/o un hilo de ejecución y un componente puede estar localizado en un ordenador y/o distribuido entre dos o más ordenadores.

Además, la expresión entrada de base de datos puede referirse a datos u otra entrada basada en el usuario que es enviada a una o más bases de datos. Una entrada de base de datos puede estar localizada en un correspondiente campo de base de datos en la base de datos. La expresión “campo de formulario” se refiere a campos que pueden aceptar la entrada de un usuario y pueden presentar también una salida; así, puede haber diferentes tipos de entrada y diferentes tipos de salida. Diferentes tipos de salida pueden corresponder a diferentes tipos de campo. Según se usa en el presente documento, la expresión “campo de entrada” también puede referirse a un campo en el cual un usuario introduce algunos datos o alguna información. Dependiendo del contexto, los campos de entrada puede referirse también campos del formulario.

La presente invención puede incorporar diversos esquemas de inferencia y/o técnicas en conexión con la generación de datos de entrenamiento para la cumplimentación automática de campos de formulario aprendida por la máquina. Según se usa en el presente documento, el término “inferencia” se refiere generalmente al proceso de razonamiento o a la deducción en cuanto a los estados del sistema, el entorno y/o el usuario a partir de un conjunto de observaciones, tal como son capturadas por medio de eventos y/o de datos. Puede emplearse la inferencia para identificar un contexto o una acción específicos, o puede generar, por ejemplo, una distribución de probabilidad sobre estados. La inferencia puede ser probabilística —es decir, el cálculo de una distribución de probabilidad sobre estados de interés en base a una consideración de los datos y los eventos—. La inferencia también puede referirse a técnicas empleadas para componer eventos de orden mayor a partir de un conjunto de eventos y/o de datos. Tal inferencia da como resultado la construcción de nuevos eventos y nuevas acciones a partir de un conjunto de eventos observados y/o de datos de eventos almacenados, estén correlacionados o no los eventos en estrecha proximidad temporal, y provengan los eventos y los datos de una o de varias fuentes de eventos y de datos.

La presente invención versa acerca del aprendizaje, para cada campo en un formulario basado en la red, de si es cumplimentable o no, por ejemplo, a partir de una o más bases de datos. Usando sistemas y técnicas de aprendizaje de máquina, los sistemas y los procedimientos de cumplimentación automática de la presente invención pueden aprender qué datos corresponden a qué campos del formulario y, entonces, cumplimentar automáticamente los campos del formulario cuando un usuario lo solicite.

Hay muchos tipos diferentes de formularios web HTML o, más bien, campos de formulario, como pueden ser denominados en el presente documento. Pueden ir de cuadros de texto, pasando por cuadros de selección, a botones de radio. Para cumplimentar con éxito automáticamente los campos de un formulario, puede ser importante determinar qué datos van en un campo particular del formulario. Típicamente, la codificación puede revelar lo que es el formulario, el lugar en el que comienza el formulario y el lugar en el que el formulario acaba. Cada campo tiene un tipo, por ejemplo texto, y puede ser nombrado con fines identificativos. Así, hay varias características asociadas con los formularios web que pueden ser aprendidas para facilitar el reconocimiento del tipo de campo del formulario y qué datos pueden o deberían ser introducidos sin el esfuerzo del usuario.

Por ejemplo, el hecho de que un campo sea un campo de texto puede ser una característica. Si el nombre del campo es “apellidos”, entonces eso indica que el contenido debería ser los apellidos de una persona, lo que puede ser una característica. El texto que rodea un campo de un formulario también importa, dado que el texto puede decir algo sobre lo que va en el campo. Así, “NombreDePila” o “Apellidos” o “Nombre” pueden indicar que en el campo deberían introducirse, respectivamente, el nombre de pila de un usuario, sus apellidos y su nombre completo.

Los botones de radio son circuitos que tradicionalmente se disponen en una lista y que pueden ser marcados haciendo clic sobre ellos. Solo un botón en un grupo de botones puede ser marcado en un momento dado. Así, marcar un botón desmarca el botón previamente marcado del grupo. Los botones de radio tienen un tipo de "radio" y los que están dispuestos en un grupo deberían tener el mismo nombre para que puedan funcionar conjuntamente. Sin embargo, cada botón dentro del grupo puede tener diferentes identificadores de texto para distinguirlos de los demás. El identificador de texto correspondiente a un botón marcado puede ser aprendido y recordado.

En la práctica, los botones de radio pueden usarse para indicar la preferencia del usuario por una forma de transporte de mercancías. Para mitigar el esfuerzo del usuario, puede cumplimentarse automáticamente el botón de radio deseado correspondiente al transporte por carretera, por ejemplo, cada vez que un usuario visita una página web relacionada con una . Las casillas de control pueden ser aprendidas de manera similar, aunque son fundamentalmente diferentes de los botones de radio, porque un usuario puede seleccionar más de una casilla de control.

Dado que la cumplimentación automática de los botones de radio y de las casillas de control no es convencional, tal cumplimentación automática puede ser inesperada por el usuario. Para abordar esto, pueden ser necesarios umbrales superiores en términos de aprendizaje y/o de para proporcionar mayores garantías de que los datos cumplimentados automáticamente son correctos.

Los cuadros de selección también pueden incluir datos que pueden ser aprendidos y cumplimentados automáticamente. Son relativamente comunes en la mayoría de los formularios web, como para proporcionar una enumeración de provincias o abreviaturas. Tal enumeración puede ser detectada, y también puede ser aprendida la enumeración o provincia particular elegida. Además de los cuadros de selección, pueden usarse botones de envío en el proceso de aprendizaje. En particular, pueden estar relacionados en el contexto con algo diferente que se mecanografió. Por ejemplo, un botón de envío llamado "adquirir" o "comprar" puede estar relacionado al contexto de lo mecanografiado en una dirección. Así, cuando se cumplimentan automáticamente formularios web nuevos o nunca vistos, la detección de que el usuario seleccionó un botón de envío llamado "comprar" puede indicar que, probablemente, también se introdujeron datos de dirección y/o de nombre.

También pueden emplearse otros indicadores para facilitar la detección de qué datos corresponden a qué campos. Por ejemplo, la anchura del campo puede ser importante. El campo "nombredepila" puede ser más corto que el campo "apellidos". De manera similar, un campo de dirección puede ser más largo que un campo de nombre. Otros matices relativos a la posición, a la disposición, el orden y similares de los campos del formulario pueden contribuir a los procesos de aprendizaje y de cumplimentación automática. Además, el aprendizaje de máquina puede ser muy sensible a estos sutiles indicios o pistas. De hecho, tales sutilezas pueden ser importantes para aprender de forma efectiva y fructífera qué datos corresponden a qué campos y las asignados de datos a los campos cuando hay presentes múltiples variaciones de los campos.

Los campos de formulario pueden ser creados para muchos tipos de diferentes de información. Además de los tipos estándar de campos de entrada como, por ejemplo, nombre de pila, apellidos, nombre completo, dirección, empresa, población, provincia, código postal, número de teléfono, dirección de correo electrónico, párrafos de texto como información de garantía y de descargo de responsabilidad, preguntas, opciones, opiniones, información personal y otros textos generales, pueden crearse campos de formulario para introducir el nombre de un fichero, típicamente cumplimentado por los usuarios navegando para encontrar el fichero. A menudo, el fichero que un usuario desea subir (por ejemplo, introduciéndolo en un campo) es el que usó más recientemente. Por lo tanto, estos campos pueden ser cumplimentados automáticamente usando ficheros a los que se ha accedido recientemente en el sistema del usuario.

En particular, algunos sistemas operativos pueden incluir una carpeta "mis documentos recientes", y este campo puede ser cumplimentado, al menos en parte, usando los ficheros usados de manera más reciente. También puede considerarse el lapso de tiempo desde que el usuario accedió a estos documentos cuando se determina si deben usarse para la cumplimentación automática. Esta reciencia puede compararse con la reciencia de otros ficheros introducidos en los campos de fichero, y puede introducirse alguna combinación de las entradas usadas recientemente y de los ficheros a los que se ha accedido recientemente en estos campos en base, al menos en parte, en los momentos del último acceso y los momentos de la última entrada.

En la práctica, imaginemos que un usuario ha modificado recientemente una foto digital y que ahora querría compartirla con la familia y los amigos subiéndola a un sitio de fotos. El campo de fichero del formulario puede ser cumplimentado automáticamente en parte examinando los ficheros a los que se ha accedido recientemente para asistir en la determinación de qué fichero introducir.

Con referencia ahora a la Fig. 1, se muestra un diagrama de bloques de alto nivel de un sistema inteligente 100 de cumplimentación automática que facilita la introducción automática de datos en campos de formulario en una página web. El sistema 100 comprende un componente 110 de aprendizaje de máquina que puede aprender qué entrada, como nombre (nombre de pila y apellidos), dirección (dirección de correo electrónico, particular, comercial, de facturación o de envío), número de teléfono (número particular, comercial, fax o móvil), código postal, provincia, nombre de fichero y/o cualesquiera otros datos, corresponde a qué campos en cualquier formulario web dado. El

componente 110 de aprendizaje de máquina puede aprender a predecir un valor de salida dado un conjunto de entradas —también denominadas datos de entrenamiento—.

Los datos de entrenamiento, que consisten en características de entrada y salida correctas, pueden ser reunidos de alguna manera. Por ejemplo, una barra de herramientas especialmente aumentada puede agregar esta información.

5 Estos datos pueden ser dados como entrada al algoritmo de entrenamiento. El algoritmo puede ser proporcionado por una cualquiera de la miríada de técnicas de aprendizaje de máquina que crean una red neural, regresión logística, modelo de máxima entropía, árbol de decisiones, modelo bayesiano ingenuo, preceptrón, Winnow, cualquier separador lineal, máquinas vectoriales de apoyo, etc. La mayoría de los tipos modelo dan salida a puntuaciones o probabilidades de cada salida posible, dados los valores de las características.

10 De las mismas resulta un modelo aprendido, que puede ser aplicado entonces a subsiguientes formularios web por medio de un componente 120 de cumplimentación automática selectiva. Cuando es invocado por un usuario, el componente 120 de cumplimentación automática selectiva puede introducir la entrada más apropiada en el campo en base, en parte, a las correlaciones aprendidas entre los datos y los campos. El componente 120 de cumplimentación automática selectiva puede ser invocado campo por campo, en grupos de campos, o de una vez  
15 para que cumplimente todos los campos cumplimentables. También los menús desplegables para cada campo pueden ser poblados como alternativa a la introducción de los datos directamente en el campo del formulario. Los elementos presentados en los menús desplegables pueden ser seleccionados en base a las correlaciones aprendidas que incluyen las entradas observadas.

20 Además, el componente 110 de aprendizaje de máquina, así como el componente 120 de cumplimentación automática selectiva puede estar ligado a una o más bases de datos, como una agenda de direcciones, una lista de contactos, un registro de cuentas, un perfil de usuario y similares. El sistema puede determinar automáticamente qué entrada de la base de datos usar en base a la información ya introducida. En particular, el componente de aprendizaje de máquina puede ser entrenado con los datos almacenados en una o más de las bases de datos. Durante el uso, el componente 120 de cumplimentación automática selectiva puede extraer datos de las respectivas  
25 bases de datos según se necesite. Por ejemplo, un usuario puede mecanografiar el nombre de pila y los apellidos de su madre y luego invocar el componente 120 de cumplimentación automática. Dado que el nombre de pila y los apellidos de la madre son introducidos manualmente por el usuario, el componente de cumplimentación automática puede detectar esto, buscar una entrada de la base de datos que contenga esos datos como nombre de pila y apellidos y cumplimentar entonces de forma automática el resto del formulario usando la información de la madre contenida en la base de datos.

Puede haber varios enfoques diferentes al empleo del aprendizaje de máquina para realizar la introducción de datos en formularios web de manera automatizada. Las Figuras 2 y 3 describen al menos dos enfoques diferentes que pueden ser combinación o empleados por separado junto con técnicas de aprendizaje de máquina. En la Fig. 2, se  
35 ilustra un diagrama de bloques de un esquema 200 de entrenamiento de cumplimentación automática que está orientado a una base de datos según un aspecto de la presente invención. El esquema 200 de entrenamiento comprende un componente 210, basado en los usuarios, que graba entradas que puede observar y/o reunir datos introducidos por un usuario. Por ejemplo, un usuario puede introducir su información de correo en un formulario basado en la red cuando se abona a una revista en línea. Pueden recogerse datos particulares y tipos de datos (por ejemplo, de texto, de casillas de control, etc.) y sus correspondientes campos de formulario (nombre, dirección de correo electrónico, dirección de facturación, etc.) y enviarse a un repositorio central 220.

40 De manera alternativa o adicional, puede pedirse a los usuarios que pueblen de antemano una o más bases de datos; por ejemplo, una para un perfil personal de usuario, una lista de contactos, una agenda de direcciones, un registro de productos y/o un registro de cuentas. Estas bases de datos pueden ser agregadas o simplemente suministradas también al repositorio central 220. Los datos o al menos una porción de los mismos que se mantiene en el repositorio central 220 pueden ser usados como entrada para un algoritmo de aprendizaje de máquina para generar uno o más conjuntos de datos de entrenamiento. Un sistema de cumplimentación automática puede ser entonces entrenado usando los datos de entrenamiento para aprender qué entradas de la base de datos  
45 cumplimentan qué campos del formulario.

Más específicamente, un componente 230 de correspondencia de entradas puede hacer uso de los datos del repositorio central estableciendo una correspondencia de datos reales con el campo apropiado del formulario. Por ejemplo, el componente 230 de correspondencia puede establecer una correspondencia del dato "98052" con un campo de un código postal estadounidense del formulario, y el hecho de que el dato numérico "98052" corresponde al campo de código postal del formulario puede ser aprendido. Así, cuando vuelva a verse el campo de código postal de formulario en un formulario web, el sistema de cumplimentación automática puede introducir o sugerir introducir  
50 "98052" en ese campo.

En situaciones en las que la privacidad es motivo de inquietud, no es preciso que los datos reales (por ejemplo, "Fred") sean enviados ni almacenados en el repositorio central 220. En vez de ello, puede enviarse el término "nombredepila" para representar los datos del usuario en el campo NOMBRE DE PILA.

En la práctica, imaginemos que un usuario está usando un navegador y se detiene en una página web de comercio electrónico como Amazon.com para comprar un libro. El usuario introduce su nombre de pila en el campo del formulario identificado como NOMBREDEPILA y sus apellidos en el campo identificado como APELLIDOS. Esta información, incluyendo la página web, puede ser recogida por una barra de herramientas instrumentada y ser almacenada en una base de datos central y luego aprendida para que, cuando el usuario o cualquier otro usuario, vuelva a visitar Amazon.com para comprar más libros, el sistema de cumplimentación automática pueda reconocer que es la misma página y luego predecir que debería introducirse, en consecuencia, la misma información (por ejemplo, el nombre de pila en el campo NOMBREDEPILA y los apellidos en el campo APELLIDOS). En general, una vez que el sistema ve y reconoce algo para lo que fue entrenado, hay una probabilidad relativamente elevada de que hará la predicción acertada la próxima vez.

Sin embargo, un enfoque alternativo que puede ser más simple de implementar es, sencillamente, para aprender, para un usuario dado, qué es lo más probable que introduzca en el campo. Este enfoque puede no usar una base de datos en absoluto. Esto puede denominarse enfoque de las entradas observadas. Por ejemplo, el enfoque de las entradas observadas puede aprender que el usuario típicamente introduce “Joshua” en un campo llamado NOMBREDEPILA en vez de aprender que habitualmente introduce el campo “nombre de pila” de su base de datos en un campo llamado NOMBREDEPILA. Un sistema de aprendizaje de mano puede ser usado aun para esto. Las salidas potenciales del sistema son todas valores de campo que un usuario ha introducido en algún momento. Por ejemplo, si ha introducido su nombre de pila, sus apellidos, varios números de teléfono y varias direcciones de correo electrónico, todos ellos se convierten en entradas potenciales. Las entradas al sistema son los mismos tipos de características que se introducen en un sistema booleano, presentado más abajo. Opcionalmente, este aprendizaje no se realiza si los campos parecen contener o relacionarse con información sensible (por ejemplo, información de tarjetas de crédito o contraseñas).

La Fig. 3 ilustra un diagrama de bloques de un esquema 300 de entrenamiento de entradas observadas que puede facilitar la cumplimentación automática de campos de formularios. El esquema 300 de entrenamiento de entradas observadas comprende un componente 310 de seguimiento de entradas que puede monitorizar y hacer seguimiento de entradas de datos realizadas por un usuario. Puesto que no se mueven datos a un repositorio de almacenamiento a largo plazo ni base de datos con fines de aprendizaje, la entrada de información puede ser más importante y/o influir más para la salida de la cumplimentación automática.

Tanto pronto como el esquema 300 detecta los datos introducidos, el componente 320 de aprendizaje puede aprender varias características sobre los datos y el formulario web. En un caso, los datos reales pueden ser asociados con el campo particular del formulario. Por ejemplo, un usuario mecanografía su nombre de pila “Joshua” en un campo NOMBRE DE PILA. Esto puede clasificarse como una característica que debe aprenderse: que la cadena de caracteres o de texto J-o-s-h-u-a corresponde a un campo NOMBRE DE PILA. Otras características incluyen observar y aprender que los campos del formulario están situados en una página web particular; que el campo NOMBRE DE PILA puede estar a la izquierda y ser más corto en anchura que el campo APELLIDOS; que si el usuario comienza a introducir datos de izquierda a derecha, entonces el campo de más a la izquierda es, probablemente, un campo de nombre de pila; que lo más probable es que los datos numéricos de más de cinco dígitos sean un número de teléfono; y diversas otras características relativas a los campos de formularios e incluso la propia página web.

Además, el esquema 300 de entradas observadas puede hacer uso a menudo de la información de orden. Por ejemplo, inmediatamente después de introducir “Joshua”, el usuario introduce a menudo “Goodman”. Inmediatamente después de introducir su prefijo local, el usuario introduce típicamente su central telefónica. Así pueden crearse características para cada valor que precede al siguiente valor.

El componente 320 de aprendizaje puede llevar a cabo un aprendizaje desconectado o conectado a la línea. En el aprendizaje sin conexión a la línea, puede almacenarse una pluralidad de casos observados de entradas de campos, incluyendo información sensible, y usarse luego como entrada para un sistema de aprendizaje de máquina en un momento posterior. Puede emplearse cualquier sistema de aprendizaje de máquina; sin embargo, el aprendizaje sin conexión a la línea tiende a llevar más tiempo que el aprendizaje con conexión a la línea. Además, dado que puede almacenarse información sensible, el aprendizaje sin conexión a la línea puede tener a ser más efectivo que el aprendizaje con conexión a la línea. El aprendizaje con conexión a la línea emplea un algoritmo de aprendizaje con conexión a la línea, como uno de los algoritmos de enfoque bayesiano ingenuo, de descenso de gradiente, Winnow o perceptrón, entre otros, que no precisa almacenar información alguna, salvo, por ejemplo, el propio modelo, que puede contener alguna información personalmente identificable. Cuando se observa un nuevo caso, el sistema 300 de cumplimentación automática puede adaptarse. En esencia, los datos introducidos por un usuario pueden ser correlacionados con el campo del formulario (por medio de un componente 330 de correspondencia entre entrada y campo) sin emplear una base de datos ni un repositorio central.

Sin embargo, los procedimientos de entradas observadas pueden ser usados en conjunción con procedimientos de entradas de bases de datos. Por ejemplo, para cualquier campo dado, puede aprenderse la probabilidad de que sea cumplimentado con una entrada de la base de datos multiplicada por una probabilidad de que el campo de la base de datos sea una entrada correcta. También puede aprenderse la probabilidad de que un campo sea cumplimentado

con una entrada que no está en la base de datos multiplicada por una probabilidad dada por el modelo de entradas observadas. Las suposiciones generadas pueden disponerse u ordenarse en base a esta información.

5 Con independencia de qué enfoque se tome, pueden generalizarse datos de entrenamiento que puedan ser útiles para muchos tipos de páginas web (por ejemplo, sitios de comercio electrónico) que pueden alterar sus páginas ligeramente de vez en cuando o presentarlas en idiomas diferentes. Los datos de entrenamiento pueden aún así producir respuestas correctas a pesar de tales cambios, que aumentan su versatilidad con respecto a llevar a cabo operaciones de cumplimentación automática con respecto a formularios basados en red.

Pasando ahora a la Fig. 4, se ilustra un formulario web HTML ejemplar 400 tomado de una página web de Verizon Wireless. Lo que sigue es un ejemplo del código HTML original:

```

10 <!--==== La tabla de información de clientes empieza aquí =====>
<TABLE WIDTH="585" BORDER="0" CELLSPACING="0" CELLPADDING="1">
  <TR>
    <TD WIDTH="5" BGCOLOR="#990000">
      <IMG SRC="https://www.verizonwireless.com/images/shared/pixel/trans.gif"
15     WIDTH="5" HEIGHT="1" BORDER="0" ALT="">
    </TD>
    <TD MIDTH="583" CLASS="smtxt3" BGCOLOR="#990000" COLSPAN="5">
      <STRONG>User Information</STRONG>
    </TD>
  </TR>
  <TR>
    <TD WIDTH="5">
      <IMG SRC="https://www.verizonwireless.com/images/shared/pixel/trans.gif"
25     Width="5" HEIGHT="1" BORDER="0" ALT="">
    </TD>
  </TR>
  <TR>
    <TD WIDTH="5">
      <IMG SRC="https://www.verizonwireless.com/images/shared/pixel/trans.gif"
30     WIDTH="5" HEIGHT="1" BORDER="0" ALT="">
    </TD>
    <TD width="93">Prefix</TD>
    <TD width="170"><STRONG><SPAN CLASS="text2">*</SPAN></STRONG>First Name</TD>
    <TD width="30">M.I.</TD>
35   <TD width="195"><STRONG><SPAN CLASS="text2">*</SPAN></STRONG>Last Name</TD>
    <TD width="186">Suffix</TD>
  </TR>
</TABLE>
40 <TABLE WIDTH="585" BORDER="0" CELLSPACING="0" CELLPADDING="1">
  <TR>
    <TD WIDTH="5">
      <IMG SRC="https://www.verizonwireless.com/images/shared/pixel/trans.gif"
45     WIDTH="5" HEIGHT="1" BORDER="0" ALT="">
    </TD>
    <TD WIDTH="93">
      <SELECT NAME="p-prefix">
        <OPTION VALUE></OPTION>
        <OPTION VALUE="MR">MR
50     <OPTION VALUE="MS">MS
        <OPTION VALUE="MRS">MRS
        <OPTION VALUE="DR">DR
      </SELECT>
    </TD>
    <TD WIDTH="135">
55   <INPUT TYPE="text" CLASS="input" NAME="p_first_name" VALUE="" SIZE="15"
      MAXLENGTH="15">
    </TD>
    <TD WIDTH="30">
60   <INPUT TYPE="text" CLASS="input" NAME="p_mid_init" VALUE="" SIZE="2"
      HAXLENGTH="1">
    </TD>
    <TD WIDTH="160">

```

## ES 2 371 918 T3

```

        <INPUT TYPE="text" CLASS="input" NAME="p_last_name" VALUE="" SIZE="18"
        MAXLENGTH="20">
    </TD>
    <TD WIDTH="184">
5      <SELECT NAME="p_suffix">
        <OPTION></OPTION>
        <OPTION VALUE="JR">JR
        <OPTION VALUE="SR">SR
10     <OPTION VALUE="2">2
        <OPTION VALUE="3">3
        <OPTION VALUE="4">4
        </SELECT>
    </TD>
    </TR>
15 </TABLE>

```

La clave para entender este código HTML es encontrar las regiones etiquetadas como "input" [entrada] (en negrita) en lo anterior. La primera entrada es el nombre de pila del usuario. Consiste en lo siguiente:

```

    <TD width="135"><INPUT TYPE="text" CLASS="input" NAME="p_first_name" VALUE=""
    SIZE="15" MAXLENGTH="15"></TD>
20 Esto dice:

```

<code>&lt;TD WIDTH="135"&gt;</code>	Una entrada de tabla de 135 píxeles de anchura.
<code>&lt;INPUT</code>	El único elemento de la entrada de tabla es un campo de entrada.
<code>TYPE="text"</code>	La entrada del usuario será texto.
<code>CLASS="input"</code>	Afecta al aspecto de la caja en algunos navegadores, y probablemente pueda ser ignorado.
<code>NAME="p_first_name"</code>	El "nombre del campo": Este es el nombre del campo tal como volverá a ser comunicado al servidor cuando se presente el formulario, como un nombre de variable en un programa de ordenador, pero en este caso se usa para la comunicación entre una página web en un cliente y en un servidor.
<code>VALUE=""</code>	El valor inicial para el campo; en este caso, la cadena vacía.
<code>SIZE="15"</code>	La anchura del campo tal como es presentado (por ejemplo, 15 caracteres de ancho).
<code>MAXLENGTH="15"&gt;</code>	La longitud máxima para la entrada en el campo.

Toda esta información puede resultar muy útil para que un sistema de aprendizaje de máquina prediga que esta es una entrada de nombre de pila. Por ejemplo, el hecho de que el nombre del campo, en inglés, contenga tanto "first" como "name" (o sea, "primer" nombre, o nombre de pila) es sumamente útil. El hecho de que la anchura y el tamaño del campo sea de 15 caracteres también es útil (obsérvese, por ejemplo, que la entrada del campo apellidos (last name en inglés) tiene una anchura de 20). Además, esta entrada de campo está en forma de tabla. Si se analiza la

```

    <TD width="170"><STRONG><SPAN CLASS="text2">*</SPAN></STRONG>First Name</TD>

```

El hecho de que la entrada anterior de la tabla no sea un campo de entrada y que contenga la cadena "First Name" es muy indicativo de que este es un campo de nombre de pila. En algunos casos, la entrada de la tabla inmediatamente a la izquierda también es útil, como en la entrada de la tabla inmediatamente a la derecha.

Por ejemplo, el hecho de que el campo inmediatamente a la derecha en este caso sea también un campo de entrada y que tenga el nombre "p\_last\_name" es también muy indicativo de que este es un campo de nombre de pila. También puede ser útil usar texto cercano o subcadenas de texto cercano que no sean idénticos o no sean simplemente equivalencias exactas.

Los sistemas de aprendizaje de máquina intentan habitualmente aprender a predecir un valor de salida dado un conjunto de entradas. En este ejemplo, los posibles valores de salida podrían ser "nombre de pila", "apellidos", "provincia", "dirección (línea 1)", "dirección (línea 2)", "nº de teléfono", etc., así como "ninguno" —para un tipo de campo que no se supiera cómo cumplimentar automáticamente—. Las entradas al sistema son valores como:

```

    Maxlength=15
    Size=15
    Fieldname="p first_name"
    Fieldname-to-right="p_last_name"
    Text-above= "<TD width="170"><STRONG><SPAN
40     CLASS="text2">*</SPAN></STRONG>First Name</TD>"
45 etc.

```

Estas entradas son convertidas entonces en valores de características, típicamente booleanos, pero a veces numéricos, categóricos y/o relacionales. En este caso, los booleanos podrían incluir:

5           MaxLengthBetween15and19 = true  
               SizeBetween15and19 = true  
               SizeExactly15 = true  
               FieldNameContains\_First = true  
               FieldnameContains\_Name = true  
               FieldNameEquals\_p\_First\_Name = true  
 10           FieldNameRightcontains\_Last = true  
               FieldNameRightContains\_Name = true  
               FieldNameRightSimeBetween20and24 = true  
               TextAboveContains\_First = true  
               TextAboveContains\_Name = true

15           Hay muchos otros valores posibles de las características, la mayoría de los cuales tendrían en este caso valores falsos.

Los valores booleanos pueden ser determinados automáticamente; por ejemplo, encontrando intervalos útiles y buscando subsecuencias de caracteres comunes útiles. De forma alternativa, los valores booleanos pueden ser determinados manualmente; por ejemplo, por medio de un programador que inspeccione valores de entrada y determine los valores booleanos más útiles.

20           Pasando ahora a la Fig. 5, se ilustra un diagrama esquemático de bloques de un sistema inteligente 500 de cumplimentación automática que ha sido entrenado según un aspecto de la presente invención. El diagrama indica que se ha presentado o mostrado al usuario (510) un nuevo formulario web y que la prestación de cumplimentación automática ha sido invocada por un usuario para cumplimentar uno o más de los campos del formulario. En particular, un componente 520 de cumplimentación selectiva puede ser activado para que recurra a la introducción automática de información en el uno o más campos del formulario.

25           Cuando se ve un nuevo formulario, pueden determinarse los valores de las características para todos los campos. Estos pueden pasarse al modelo aprendido, que da salida a probabilidades, puntuaciones, o valores 0-1 para cada salida (tipo de campo) para cada campo. Por ejemplo, los campos del formulario y sus valores pueden correlacionarse: un campo puede tener un 90% de probabilidad de ser el primero y un 10% de no ser el primero (o un 10% de ser el último) y otro puede tener un 90% de probabilidad de ser el último y un 10% de ser el primero. Así, deberían encontrarse las mejores asignaciones para facilitar la cumplimentación de datos en los campos de los formularios con la mayor cantidad de precisión.

30           La asignación de entradas de datos (o entradas de base de datos) a valores del formulario puede basarse, al menos en parte, en una o más restricciones 530. Las restricciones 530 pueden establecerse para limitar o evitar que se hagan ciertos tipos de asignaciones. Un ejemplo de un límite restrictivo se retrotrae al número de veces que puede usarse una entrada de la base de datos en un formulario dado. Típicamente, una entrada de la base de datos debería usarse una sola vez por formulario, aunque, en algunos casos (por ejemplo, un número de teléfono), puede ser apropiado permitir la reutilización. Si se proscribe o se prohíbe la reutilización de ciertos campos (por ejemplo, se establece como restricción), puede encontrarse la asignación más probable, o de más puntuación, de los nombres de campos a los nombres de la base de datos con las restricciones de reutilización. Así, un componente 540 de análisis de entrada de datos a campo de formulario puede evaluar la una o más restricciones aplicables para el formulario web dado y luego determinar asignaciones de entradas de datos (desde una base de datos opcional 550) a los campos del formulario. Al final, las salidas "sugeridas" del sistema de cumplimentación automática pueden ser presentadas en los respectivos campos del formulario.

35           También podría haber presentes tipos de salida contradictorios, como, en inglés, "first name", "last name" y "name" (estando compuesto este tanto del nombre de pila como de los apellidos), por ejemplo. Así, puede establecerse una restricción de composición que asuma que estas diferentes variaciones deberían existir en la misma página. La restricción puede dictar que solo puede haber un campo "name" o campos de nombres "first" y "last". Si está presente el campo "name", entonces se construyen los datos, por ejemplo, concatenando los campos de los nombres first y last de la base de datos. De manera similar, podría haber campos, en una página estadounidense, "area code" [prefijo], "exchange" [central telefónica] y "last4digits" [últimos4dígitos], así como un campo "phone number" [número de teléfono] (consistente en los 3 campos de la base de datos concatenados). En cualquier conjunto de circunstancias, también puede determinarse la asignación más probable o de más puntuación de los nombre de los campos coherente con las restricciones de composición (dado que no se permite que haya tanto una entrada para el nombre completo y una entrada para los apellidos).

45           Algunas entradas de campos nunca deberían ser usadas sin otras entradas de campos. Por ejemplo, entrada de la segunda línea (nº 2) de la dirección no podría usarse nunca sin la entrada de la primera línea (nº 1) de la dirección. La inicial central de un usuario anglosajón no podría usarse nunca sin su nombre de pila. De modo similar, la inicial central nunca podría ocurrir antes que el campo nombre de pila; la línea 2 de la dirección nunca podría ocurrir antes

que la línea 1 de la dirección. Además, también puede encontrarse la asignación más probable o de mayor puntuación de los nombres de los campos coherente con estas restricciones.

Además, si una persona está cumplimentando información de izquierda a derecha, el sistema puede adivinar que la entrada de datos inicial corresponde al nombre de pila y luego predecir que el campo a la derecha es el campo de los apellidos. Por lo tanto, pueden realizarse observaciones en consideración de las restricciones para determinar mejor las asignaciones de campos de entrada de datos a campos del formulario.

También las características exclusivas de algunos campos del formulario pueden estar limitadas a restricciones. Por ejemplo, puede haber presente una característica de que un campo particular de la base de datos esté encima, debajo, a la izquierda, a la derecha, inmediatamente debajo, inmediatamente encima, inmediatamente a la izquierda o inmediatamente a la derecha de algún otro campo. Así, en la práctica, en un formulario estadounidense, cabría esperar que el campo “exchange” estuviera a la derecha del campo “area code”. Cuando se emplea el sistema 500, puede ser posible “saber” qué campos de entrada son cuáles, pero pueden ensayarse muchas asignaciones diferentes y puede realizarse un intento de escoger uno que sea coherente con las restricciones.

De forma alternativa, pueden realizarse varias asignaciones diferentes. Partiendo de ello, pueden determinarse las probabilidades maximizadas conjuntas de asignaciones particulares de los campos de entrada a campos del formulario que incluyan estas características de diseño. Esto puede lograrse, en parte, condicionando lo siguiente: ¿Cuál es la probabilidad de que este primer elemento sea el primero y de que un segundo elemento sea el segundo dado el primer elemento, y de que el tercer elemento sea el tercero dados los elementos primero y segundo, etcétera? Cuando esto se multiplica, puede obtenerse una probabilidad global conjunta. Así, en vez de pensar en este problema en el sentido de asignar cada campo de forma independiente de los demás, la mejor asignación de los campos puede determinarse en términos de maximizar las probabilidades conjuntas.

Las restricciones 530 pueden ser restricciones inflexibles o flexibles. En una restricción inflexible, la restricción debe ser satisfecha. En una restricción flexible, hay una fuerte sanción por violar la restricción.

Gran parte de las exposiciones previas han implicado la construcción de un sistema de cumplimentación automática que funciona para cada usuario (no específico para el usuario). Sin embargo, el sistema de cumplimentación automática también puede ser personalizado según el usuario. En particular, la correspondencia de campos de la base de datos para formar valores puede ser personalizada. Por ejemplo, puede observarse que, en una página particular, un usuario rechaza la entrada automática sugerida e introduce datos manualmente. Parte de estos datos pueden coincidir con entradas en la base de datos. Así, puede aprenderse que, para este usuario, estas entradas de campos de entrada corresponden a las entradas observadas de la base de datos. Acto seguido, el cumplimentador automático del usuario puede ser actualizado con esta información. La información también puede ser actualizada en una forma específica al sitio, aprendiendo con ello que las entradas en una página web particular corresponden a entradas particulares de la base de datos. Además, el usuario puede personalizar el sistema de cumplimentación automática desde el comienzo, o pueden presentársele configuraciones por defecto que puedan ser por el usuario según se necesite.

La personalización puede llevarse a cabo de una forma general específica al sitio o aprendida por la máquina. Por ejemplo, el sistema puede observar que cuando un campo dice NOMBRE, en español, el usuario introduce “José” — no Juan, que puede ser la sugerencia de cumplimentación automática—. Si es específico al sitio, el sistema puede observar que el usuario mecanografió “José” en esta página particular. En general, las entradas de la base de datos y/o la personalización pueden almacenarse opcionalmente en un ordenador separado del ordenador del usuario (por ejemplo, en un servidor al que se accede por medio de Internet o en una LAN).

Algunos formularios, como la página de ejemplo anterior, pueden contener múltiples conjuntos de campos (por ejemplo, una dirección particular y una dirección comercial). Pueden usarse la heurística o el aprendizaje para identificar tales casos. Por ejemplo, en la página de Verizon (Fig. 4), las diferentes direcciones están en tablas diferentes. Típicamente, las diferentes direcciones pueden estar espacialmente disjuntas y/o relativamente alejadas en el texto real de la página. Así, puede aprenderse a separar la información en conjuntos respectivos (por ejemplo, la dirección particular y la dirección comercial).

En algunos casos, el sistema de cumplimentación automática puede no sugerir de inmediato a un usuario las entradas correctas de un formulario particular. Cuando un usuario introduce manualmente datos en un formulario, puede observarse lo que introduce el usuario e intentar determinarse otras correspondencias de los campos de entrada y los campos de la base de datos coherente con lo que introduce, o una entrada diferente en la base de datos. Por ejemplo, el sistema puede introducir la dirección particular del usuario, pero observar luego que introduce datos coherentes con su dirección comercial. De forma alternativa, el sistema puede observar que el usuario introduce datos coherentes con una entrada en su lista de contactos. En este caso, pueden crearse nuevas entradas automáticas de campos en base a las entradas coherentes mejor puntuadas.

Las sugerencias para las entradas pueden realizarse usando una interfaz de usuario de “todo de golpe”, como se hace en la actualidad, pero también puede hacerse de forma incremental (por ejemplo, a través de una lista desplegable en cada campo), con o sin una base de datos de campos y con o sin aprendizaje de máquina. Cuando

el usuario comienza a introducir datos en un campo, puede haber un cuadro desplegable con las mejores sugerencias basadas en parte en lo que el usuario está introduciendo (por ejemplo, el usuario mecanografía "J" y un cuadro desplegable ofrece "Joshua"). Alternativamente, en el momento en que el usuario entra en un campo de entrada (por ejemplo, usando las teclas de flechas, pulsando el tabulador o haciendo clic en el campo), el campo puede ser rellenado con la mejor sugerencia. Cuando un usuario mecanografía en un campo que contiene texto seleccionado, el texto seleccionado puede ser eliminado automáticamente; pero si el usuario no mecanografía nada, el texto puede permanecer ahí.

Ahora se describirán diversas metodologías según la presente invención por medio de una serie de acciones. Debe entenderse y apreciarse que la presente invención no está limitada por el orden de las acciones, ya que algunas acciones pueden ocurrir, según la presente invención, en órdenes diferentes y/o simultáneamente con otras acciones de lo mostrado y descrito en el presente documento. Por ejemplo, los expertos en la técnica comprenderán y apreciarán que una metodología podría ser representada de forma alternativa como una serie de estados o eventos interrelacionados, como en un diagrama de estado. Además, no todas las acciones ilustradas pueden ser necesarias para implementar una metodología según la presente invención.

Con referencia ahora a la Fig. 6, se muestra un diagrama de flujo de un procedimiento ejemplar 600 que facilita la entrada automática de datos en uno o más campos de un formulario en una página web según un aspecto de la presente invención. El procedimiento 600 implica proporcionar o generar un formulario basado en la red en 610. El formulario basado en la red puede incluir uno o más campos de formulario que un usuario deseara cumplimentar de manera automatizada. Esto minimiza los errores de usuario y puede mitigar los de entrada de datos. En 620, puede establecerse una correspondencia entre la entrada de datos y un respectivo campo de entrada de datos usando una técnica de aprendizaje de máquina. En consecuencia, el sistema de aprendizaje de máquina puede aprender qué datos corresponden a qué campos.

Con referencia a la Fig. 7, se muestra un diagrama de flujo de un procedimiento ejemplar 700 que facilita la generación de datos de entrenamiento para su uso en un sistema de aprendizaje de máquina en conexión con la construcción de un sistema inteligente de cumplimentación automática. El procedimiento 700 puede comenzar con un usuario que introduce datos en uno o más campos de formulario en una página web en 710. En 720, uno o más de los elementos siguientes puede observarse y/o reunirse por medio de una o más barras de herramientas instrumentadas: datos introducidos en al menos un respectivo campo del formulario; qué entrada de datos va a qué campo del formulario; y/o características de los respectivos campos del formulario. En 730, pueden emplearse uno o más sistemas de aprendizaje de máquina para aprender las correlaciones entre las entradas de datos y el campo correcto del formulario, de modo que cuando se visita una página nueva o se vuelve a visitar una página previa, el procedimiento 70 pueda saber qué datos introducir en qué campos de manera automatizada. Tal aprendizaje puede implicar aprendizaje sin conexión a la línea o aprendizaje con conexión a la línea.

Opcionalmente, los datos recogidos en 720 pueden ser devueltos al repositorio central o a la base de datos en 810, como se demuestra en la Fig. 8. Los datos del repositorio central puede ser correlacionados con sus campos respectivos, resultando en la generación de uno o más conjuntos de datos de entrenamiento. Los datos de entrenamiento pueden usarse entonces como entradas al sistema de aprendizaje de máquina para aprender qué campos o entradas de la base de datos corresponden a qué campos del formulario. Como resultado de tal aprendizaje, los campos de formulario en una página web del usuario pueden ser cumplimentados automáticamente con independencia de si la página web ha sido vista antes o no.

Pasando ahora a la Fig. 9, se muestra un diagrama de flujo de un procedimiento ejemplar 900 que facilita la asignación de entradas de base de datos a campos de formulario, particularmente cuando el formulario web no ha sido visto nunca antes por el sistema o el procedimiento de cumplimentación automática. En tales circunstancias, el procedimiento 900 puede encontrar inicialmente valores de características para los campos del formulario en 910. Los valores de las características pueden ser probabilidades o puntuaciones, dependiendo de cómo se calculen. En 920, pueden establecerse una o más restricciones para regular o controlar las asignaciones de las entradas de datos (o entradas de la base de datos) a campos del formulario. Hay varios tipos de restricciones que pueden emplearse, tales como, sin limitación, restricciones de composición (por ejemplo, tipos de salida contradictorios), restricciones de reutilización (por ejemplo, uso de la misma entrada de datos más de una vez por formulario), restricciones de orden (por ejemplo, algunos formularios están situados necesariamente en un cierto orden, una cierta posición o una cierta distancia de otros campos del formulario) y/o restricciones de acompañamiento (por ejemplo, algunos formularios no pueden usarse independientemente de otros formularios).

En 930, las asignaciones de entradas de datos a campos del formulario puede determinarse en base, al menos en parte, a una o más restricciones. Las restricciones pueden incluir restricciones tanto inflexibles como flexibles. A diferencia de las restricciones inflexibles, no es preciso que las restricciones flexibles sean satisfechas. Sin embargo, no satisfacerlas puede llevar a una fuerte sanción. En vez de determinar la asignación de cada campo del formulario independientemente de los otros campos del formulario, ya sean próximos o no, en 940 puede calcularse una probabilidad conjunta maximizada para determinar la probabilidad global de las asignaciones. Pueden crearse muchos conjuntos diferentes de asignaciones, y las respectivas probabilidades de cada asignación dentro de un conjunto de asignaciones dadas las otras asignaciones dentro de ese conjunto pueden multiplicarse entre sí. Puede

seleccionarse y utilizarse el conjunto de asignaciones que tenga la mayor probabilidad conjunta para cumplimentar automáticamente un formulario basado en la red.

Con referencia ahora a la Fig. 10, se muestra un diagrama de flujo de un procedimiento ejemplar 1000 que demuestra la invocación de un procedimiento o un sistema inteligentes de cumplimentación automática según se describe más arriba. En este escenario particular, pueden construirse una o más bases de datos en 1010 y los datos almacenados en las mismas pueden ser correlacionados para casar una entrada de la base de datos con un campo del formulario. En 1020, imaginemos que un usuario ha abierto una página de un formulario basado en la red que tiene uno o más campos para cumplimentar los datos del usuario. En 1030, el usuario invoca selectivamente una función de cumplimentación automática por campo o por grupos de campos.

Por ejemplo, la característica de cumplimentación automática puede ser invocada de una sola vez para cumplimentar todos los campos de golpe, campo por campo, con lo cual cada campo puede ser cumplimentado automáticamente según desee el usuario y/o por grupo de campos (por ejemplo, los campos que están relacionados, como el campo de dirección, el campo de población y el campo de provincia). Suponiendo que el al menos susodicho campo ha sido cumplimentado automáticamente, el usuario puede invalidar los datos sencillamente mecanografiando sobre ellos. Estos "nuevos" datos pueden ser aprendidos también y el componente de cumplimentación automática puede ser modificado para reconocer cuándo deberían usarse estos nuevos datos. Tras la invalidación, la características de cumplimentación automática puede ser invocada de nuevo, por ejemplo para cumplimentar los campos restantes. Sin embargo, los campos sobrescritos no se ven afectados. Además, la característica de cumplimentación automática puede observar los datos introducidos manualmente por el usuario y luego cumplimentar automáticamente el resto de los campos usando datos que sean relevantes a los datos sobrescritos.

Por ejemplo, imaginemos que un campo NOMBRE es cumplimentado de forma automática inicialmente. El usuario borra su nombre e inserta el nombre de su hermano. La información de su hermano puede que esté almacenada en una de las bases de datos accesibles por el componente de cumplimentación automática. Así, tras reconocer el nombre del hermano, el componente o el procedimiento de cumplimentación automática pueden proceder a sacar los datos del hermano de la base de datos y poblar en consecuencia los campos del formulario.

Para proporcionar un contexto adicional para diversos aspectos de la presente invención, la Fig. 11 y la siguiente exposición están pensadas para proporcionar una breve descripción general de un entorno operativo adecuado 1110 en el que diversos aspectos de la presente invención pueden ser implementados. Aunque la invención está descrita en el contexto general de instrucciones ejecutables por ordenador, como módulos de programa ejecutados por uno o más ordenadores u otros dispositivos, los expertos en la técnica reconocerán que la invención también puede ser implementada en combinación con otros módulos de programa y/o como una combinación de soporte físico y soporte lógico.

Sin embargo, en general, los módulos de programa incluyen rutinas, programas, objetos, componentes, estructuras de datos, etc. que llevan a cabo tareas particulares o implementan tipos particulares de datos. El entorno operativo 1110 es únicamente un ejemplo de un entorno operativo y no se pretende sugerir ninguna limitación en cuanto al alcance del uso o la funcionalidad de la invención. Otros sistemas, entornos y/o configuraciones bien conocidos de ordenadores que pueden ser adecuados para su uso con la invención incluyen, sin limitación, ordenadores personales, dispositivos de mano o portátiles, sistemas multiprocesador, sistemas basados en microprocesadores, electrónica programable de consumo, PC de red, miniordenadores, ordenadores centrales, entornos informáticos distribuidos que incluyen los anteriores sistemas o dispositivos y similares.

Con referencia a la Fig. 11, un entorno ejemplar 1110 para implementar diversos aspectos de la invención incluye un ordenador 1112. El ordenador 1112 incluye una unidad 1114 de proceso, una memoria 1116 del sistema y un bus 1118 del sistema. El bus 1118 del sistema acopla componentes del sistema, incluyendo, sin limitación, la memoria del sistema 1116, a la unidad 1114 de proceso. La unidad 1114 de proceso puede ser cualquiera de diversos procesadores disponibles. También pueden emplearse microprocesadores duales y otras arquitecturas de procesadores múltiples como la unidad 1114 de proceso.

El bus 1118 del sistema puede ser de cualquier de varios tipos de estructuras de bus, incluyendo el bus de memoria o el controlador de memoria, un bus de periféricos o un bus externo y/o un bus local que usa cualquier variedad de las arquitecturas de bus disponibles, incluyendo, sin limitación, un bus de 11 bits, la Arquitectura Industrial Normalizada (ISA), la Arquitectura de Microcanal (MSA), la ISA Extendida (EISA), Electrónica Integrada de Dispositivos (IDE), Bus Local VESA (VLB), Interconexión de Componentes Periféricos (PCI), Bus Serie Universal (USB), Puerto de Gráficos Acelerado (AGP), bus de la Asociación Internacional de Fabricantes de Memoria para Ordenadores Personales (PCMCIA) y la Interfaz para Sistemas de Ordenadores Pequeños (SCSI).

La memoria 1116 del sistema incluye memoria volátil 1120 y memoria no volátil 1122. El sistema básico de entrada/salida (BIOS), que contiene las rutinas básicas para transferir información entre elementos dentro del ordenador 1112, tal como en el arranque, está almacenado en la memoria no volátil 1122. A título de ilustración, y no de limitación, la memoria no volátil 1122 puede incluir memoria de solo lectura (ROM), ROM programable (PROM), ROM eléctricamente programable (EPROM), ROM eléctricamente borrable (EEPROM) o memoria flash. La memoria

volátil 1120 incluye memoria de acceso directo (RAM), que actúa como memoria asociada externa. A título de ilustración y no de limitación, hay RAM disponible en muchas formas, como la RAM síncrona (SRAM), la RAM dinámica (DRAM), la DRAM síncrona (SDRAM), la SDRAM de doble tasa de datos (DDR SDRAM), la SDRAM mejorada (ESDRAM), la Synchlink DRAM (SLDRAM) y la Rambus RAM (DRRAM).

- 5 El ordenador 1112 también incluye medios de almacenamiento de ordenador extraíbles/no extraíbles volátiles y no volátiles. La Fig. 11 ilustra, por ejemplo, un almacenamiento 1124 de disco. El almacenamiento 1124 de disco incluye, sino limitación, dispositivos como una unidad de disco magnético, una unidad de disco flexible, una unidad de cinta, una unidad Jaz, una unidad Zip, una unidad LS-100, una tarjeta de memoria flash o lápiz USB. Además, el almacenamiento 1124 de disco puede incluir medios de almacenamiento por separado o en combinación con otros  
10 medios de almacenamiento, incluyendo, sin limitación, una unidad de disco óptico, como un dispositivo de disco compacto ROM (CD-ROM), una unidad de CD grabable (unidad CD-R), una unidad de CD regrabable (unidad CD-RW) o una unidad de disco versátil digital ROM (DVD-ROM). Para facilitar la conexión de los dispositivos 1124 de almacenamiento de disco al bus 1118 del sistema, típicamente se usa una interfaz extraíble o no extraíble, como la interfaz 1126.
- 15 Ha de apreciarse que la Fig. 1 describe un soporte lógico que actúa como intermediario entre usuarios y los recursos básicos del ordenador descritos en el entorno operativo adecuado 1110. Tal soporte lógico incluye un sistema operativo 1128. El sistema operativo 1128, que puede estar almacenado en el almacenamiento 1124 de disco, actúa controlando y asignando recursos del sistema 1112 de ordenador. Las aplicaciones 1130 del sistema pueden aprovecharse de la gestión de recursos por parte del sistema operativo 1128 a través de módulos 1132 de programa y datos 1134 de programa almacenados ya sea en la memoria 1116 del sistema o en el almacenamiento 1124 de disco. Ha de apreciarse que la presente invención puede ser implementada con diversos sistemas operativos o combinación de sistemas operativos.

Un usuario introduce órdenes o información en el ordenador 1112 a través de uno o varios dispositivos 1136 de entrada. Los dispositivos 1136 de entrada incluyen, sin limitación, un dispositivo de puntero como un ratón, una bola de mando, un estilete, un tablero táctil, un teclado, un micrófono, una palanca de control, un mando para juegos, una antena parabólica, una tarjeta de sintonización de TV, una cámara digital, una cámara de vídeo digital, una cámara web y similares. Estos y otros dispositivos de entrada se conectan a la unidad 1114 de proceso por medio del bus 1118 del sistema a través de uno o varios puertos 1138 de interfaz. El o los puertos 1138 de interfaz incluyen, por ejemplo, un puerto serie, un puerto paralelo, un puerto de juegos, un puerto bus serie universal (USB). El o los dispositivos 1140 de salida usando algunos de los mismos tipos de puertos que el o los dispositivos 1136 de entrada. Así, por ejemplo, puede usarse un puerto bus serie universal USB para proporcionar una entrada al ordenador 1112 y para dar salida a información del ordenador 1112 a un dispositivo 1140 de salida. Se proporciona el adaptador 1142 de salida para ilustrar que hay algunos dispositivos 1140 de salida, como monitores, altavoces e impresoras, entre otros dispositivos 1140 de salida, que requieren adaptadores especiales. Los adaptadores 1142 de salida incluyen, a título de ilustración y no de limitación, tarjetas de vídeo y de sonido que proporcionan un medio de conexión entre el dispositivo 1140 de salida y el bus 1138 del sistema. Debería observarse que otros dispositivos y/o sistemas de dispositivos proporcionan capacidades tanto de entrada como de salida, como un o varios ordenadores remotos 1144.

El ordenador 1112 puede operar en un entorno de red usando conexiones lógicas a uno o más ordenadores remotos, como el o los ordenadores remotos 1144. El o los ordenadores remotos 1144 pueden ser un ordenador personal, un servidor, un dispositivo de encaminamiento, un PC de red, una estación de trabajo, un aparato eléctrico basado en microprocesadores, un dispositivo del mismo nivel u otro nodo de red común y similares, y, típicamente, incluye muchos o todos los elementos descritos en relación con el ordenador 1112. Con fines de brevedad, solo se ilustra el dispositivo 1146 de almacenamiento con uno o varios ordenadores remotos 1144. El o los ordenadores remotos 1144 están conectados lógicamente al ordenador 1112 a través de una interfaz 1148 de red, y luego conectados físicamente por medio de una conexión 1150 de comunicaciones. La interfaz 1148 de red abarca redes de comunicaciones como redes de área local (LAN) y redes de área amplia (WAN). Las tecnologías LAN incluyen la Interfaz de Datos Distribuidos por Fibra Óptica (FDDI), la Interfaz de Datos Distribuidos por Cobre (CDDI), Ethernet/IEEE 1102.3, Token Ring/IEEE 1102.5 y similares. Las tecnologías WAN incluyen, sin limitación, enlaces punto a punto, redes de conmutación por circuitos como las Redes Digitales de Servicios Integrados (RDSI) y variaciones de las mismas, redes conmutadas por paquetes y Líneas Digitales de Abonado (DSL).

La o las conexiones 1150 de comunicaciones se refieren al soporte físico/lógico empleados para conectar la interfaz 1148 de red al bus 1118. Aunque se muestra la conexión 1150 de comunicaciones con claridad ilustrativa dentro del ordenador 1112, también puede ser externa al ordenador 1112. El soporte físico/lógico necesario para la conexión a la interfaz 1148 de red incluye, solo con fines ejemplares, tecnologías internas y externas como módems, incluyendo módems regulares de tipo telefónico, módems de cable y módems DSL, adaptadores RDSI y tarjetas Ethernet.

Lo que se ha descrito en lo que antecede incluye ejemplos de la presente invención. Por supuesto, no es posible describir toda combinación concebible de componentes o metodologías con el fin de describir la presente invención, pero una persona con un dominio normal de la técnica puede reconocer que son posibles muchas combinaciones y permutaciones adicionales de la presente invención. En consecuencia, se pretende que la presente invención

abarque todas las alternaciones, modificaciones y variaciones de ese tipo que caigan dentro del alcance de las reivindicaciones adjuntas. Además, hasta el grado en que el término "incluye" se usa, ya sea en la descripción detallada o en las reivindicaciones, se pretende que tal término sea inclusivo, de manera similar a como se interpreta el término "comprende" o "comprendiendo" cuando se emplea como palabra de transición en una reivindicación.

**REIVINDICACIONES**

1. Un sistema de cumplimentación automática que comprende:
  - un componente de aprendizaje de máquina que usa uno o más conjuntos de entradas para facilitar la cumplimentación automática de formularios basados en la red; y
  - 5 un componente de cumplimentación automática que introduce selectivamente datos en uno o más campos del formulario en base, al menos en parte, a la información generada por el componente de aprendizaje de máquina;

**caracterizado porque**

  - 10 el componente de aprendizaje de máquina comprende un componente de monitorización que observa las entradas del usuario en los campos del formulario; y
  - en el que el componente de aprendizaje de máquina hace uso de una o más entradas de una base de datos y de las entradas observadas para aprender al menos una de las siguientes:
    - una probabilidad de que un campo sea cumplimentado con una entrada de la base de datos multiplicada por una probabilidad de que el campo de la base de datos sea una entrada correcta; y
    - 15 una probabilidad de que un campo no sea cumplimentado con una entrada de la base de datos multiplicada por una probabilidad calculada observado las entradas del usuario.
2. El sistema de la reivindicación 1 en el que uno o más conjuntos de entradas se almacenan en una o más bases de datos y se recuperan de las mismas.
3. El sistema de la reivindicación 2 en el que el componente de aprendizaje de máquina aprende a establecer una correspondencia entre una pluralidad de campos de la base de datos y los correspondientes campos del formulario.
4. El sistema de la reivindicación 2, comprendiendo la una o más bases de datos al menos una base de datos de cumplimentación automática, una base de datos de un registro de cuentas y una base de datos de un registro de productos.
5. El sistema de la reivindicación 2 en el que el componente de aprendizaje de máquina genera asignaciones de entradas de la base de datos a campos del formulario en base, al menos en parte, a una o más restricciones.
6. El sistema de la reivindicación 5, comprendiendo la una o más restricciones restricciones de reutilización que limitan el número de veces que una entrada cualquiera de la base de datos es usada por formulario basado en la red.
7. El sistema de la reivindicación 5, comprendiendo la una o más restricciones restricciones que controlan qué entradas o tipos de salida de la base de datos se permite que estén presentes simultáneamente en cualquier formulario basado en la red.
8. El sistema de la reivindicación 5, comprendiendo la una o más restricciones restricciones en campos que ocurren juntos o en un orden particular.
9. El sistema de la reivindicación 2 en el que la entrada en la base de datos corresponde a uno o más campos de la base de datos o de entrada, teniendo los campos de la base de datos o de entrada una o más características derivadas de la misma.
10. El sistema de la reivindicación 9, siendo al menos una característica para un campo particular de la base de datos que está, al menos, en una de estas posiciones: encima, debajo, a la izquierda, a la derecha, inmediatamente debajo, inmediatamente encima, inmediatamente a la izquierda o inmediatamente a la derecha de algún otro campo de la base de datos.
11. El sistema de la reivindicación 10 que, además, comprende un componente que determina la probabilidad conjunto de una asignación particular de campos de entrada a campos del formulario que incluye la al menos susodicha característica.
12. El sistema de la reivindicación 2 que, además, comprende un componente que personaliza la correspondencia entre campos de la base de datos y valores del formulario aprendiendo que, para un usuario particular, algunas entradas de campos introducidas corresponden a las entradas observadas de la base de datos y que actualiza el componente de cumplimentación automática con tal información.

- 5
13. El sistema de la reivindicación 12 en el que el componente actualiza el componente de cumplimentación automática de una manera específica para el sitio que implica aprender que algunas entradas de campo introducidas en una página web particular corresponden a entradas particulares de la base de datos.
14. El sistema de la reivindicación 2 en el que el componente de aprendizaje de máquina observa que un usuario está introduciendo datos en un campo coherentes con los datos encontrados en la una o más bases de datos, lo que activa el componente de cumplimentación automática para que proporcione nuevas sugerencias de cumplimentación automática en base, en parte, en las entradas coherentes mejor clasificadas.
15. El sistema de la reivindicación 2 en el que la una o más bases de datos se mantienen en una máquina separada del ordenador del usuario.
- 10
16. El sistema de la reivindicación 1 que, además, comprende un componente de recogida de datos que reúne datos de entrenamiento de una o más fuentes.
17. El sistema de la reivindicación 16 en el que el componente de recogida de datos agrega la entrada y la envía a un repositorio central para un tratamiento ulterior.
18. El sistema de la reivindicación 16 en el que el componente de recogida de datos comprende al menos una barra de herramientas instrumentada.
- 15
19. El sistema de la reivindicación 1, comprendiendo la MAXLENGTH o SIZE.
20. El sistema de la reivindicación 1, comprendiendo la entrada subcadenas .
21. El sistema de la reivindicación 1, comprendiendo la entrada información de uno o más campos adyacentes del formulario.
- 20
22. El sistema de la reivindicación 1, comprendiendo la entrada al menos uno de un texto cercano o subcadenas de un texto cercano.
23. El sistema de la reivindicación 1 en el que el componente de aprendizaje de máquina convierte una o más entradas en uno o más valores de una característica.
- 25
24. El sistema de la reivindicación 23 en el que los valores de las características son valores booleanos o valores numéricos.
25. El sistema de la reivindicación 1 en el que el componente de cumplimentación automática proporciona sugerencia de cumplimentación automática en un cuadro desplegable desde un campo cuando un usuario empieza a introducir datos en el campo.
- 30
26. El sistema de la reivindicación 1 en el que el componente de cumplimentación automática cumplimenta un campo con datos sugeridos tan pronto como un usuario entra en el campo.
27. El sistema de la reivindicación 1 en el que el componente de aprendizaje de máquina aprende cuál de las entradas que ha realizado previamente un usuario es más probable que introduzca en un campo dado de entrada.
- 35
28. El sistema de la reivindicación 27 en el que el componente de aprendizaje de máquina no aprende entradas del usuario para campos que parecen comprender información sensible.
29. El sistema de la reivindicación 1, generando el componente de aprendizaje de máquina una o más características en base, en parte, a los posibles valores para los campos.
30. El sistema de la reivindicación 29 en el que la una o más características están basadas en el orden para cada valor que precede a un valor siguiente.
- 40
31. El sistema de la reivindicación 1 en el que el componente de aprendizaje de máquina es entrenado usando un algoritmo desconectado de la línea que incluye reunir una pluralidad de los casos observados de entradas en campos y almacenarlos.
- 45
32. El sistema de la reivindicación 1 en el que el componente de aprendizaje de máquina es entrenado usando un algoritmo en línea usando uno cualquiera de los algoritmos de enfoque bayesiano ingenuo, de descenso de gradiente, Winnow y preceptrón.
33. El sistema de la reivindicación 1 en el que los campos del formulario comprenden un campo de entrada de ficheros que presenta un fichero en base, en parte, a ficheros a los que se ha accedido recientemente.

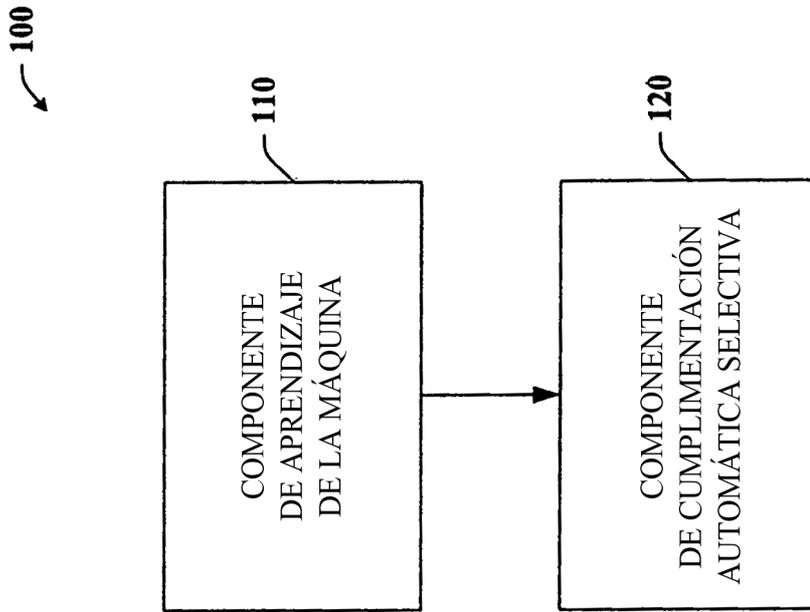
**34.** El sistema de la reivindicación 33, determinándose el fichero introducido por el componente de cumplimentación automática en base, al menos en parte, a alguna combinación de entradas usadas recientemente en el campo de ficheros y de ficheros a los que se ha accedido recientemente.

**35.** El sistema de la reivindicación 34, basándose la combinación, al menos en parte, en un tiempo de último acceso y un tiempo de la última entrada.

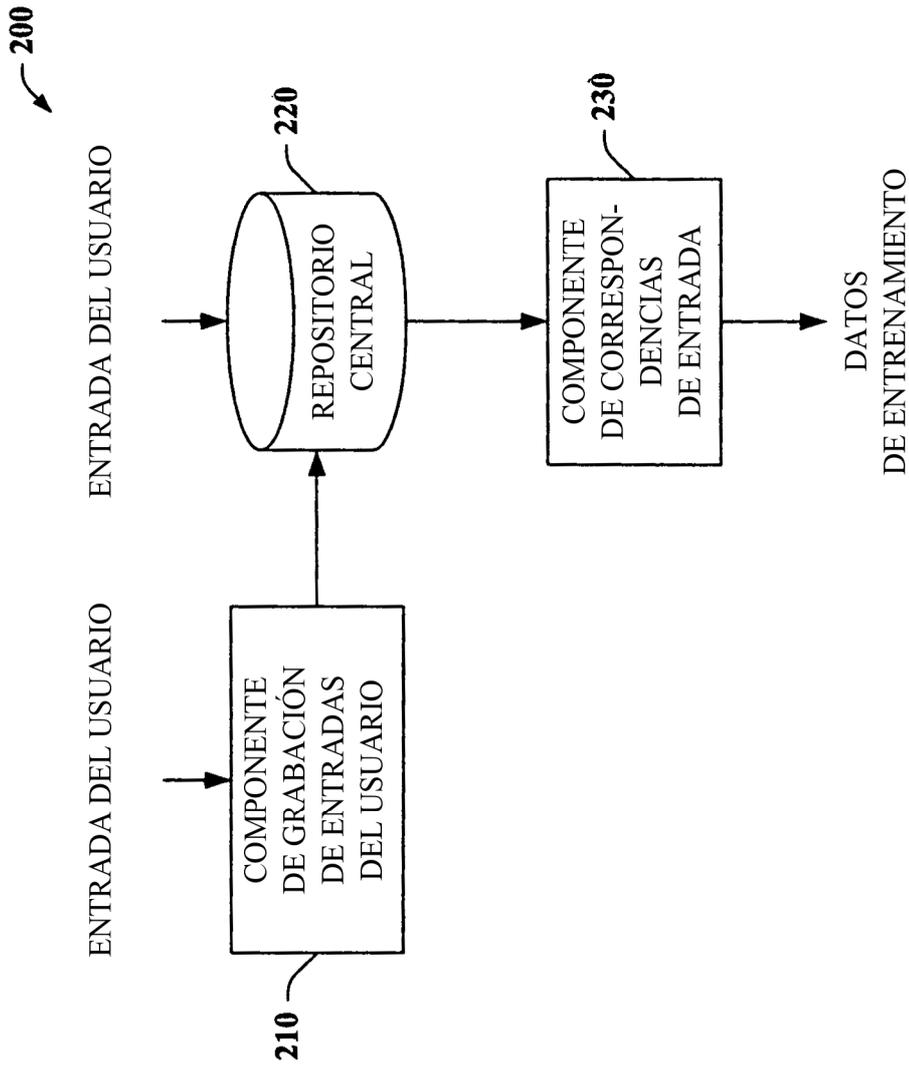
**36.** El sistema de la reivindicación 1, comprendiendo los campos del formulario botones de radio y casillas de control, de modo que el componente de cumplimentación automática pueda aprender si marcar una casilla o dejarla sin marcar o a seleccionar un botón de radio.

**37.** Un medio legible por ordenador que tiene almacenado en el mismo el sistema de la reivindicación 1.

10



**FIG. 1**



**FIG. 2**

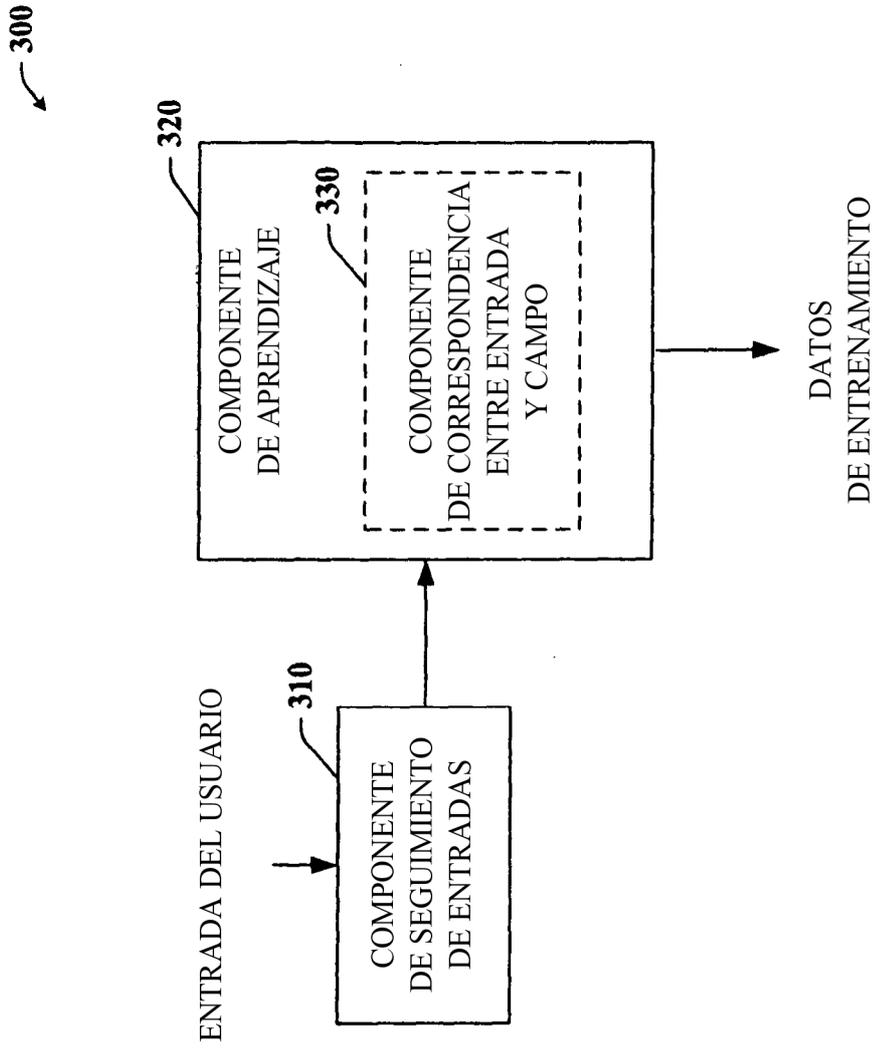


FIG. 3

400

Purchase a Plan or Phone - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media Print Mail Stop

Address C:\misc\verizoncheckout.htm Go Links

To complete your order for Verizon Wireless service, please complete the following information. All information provided will be transmitted using a secure connection to protect confidential data.

A red asterisk \* denotes required fields.

**User Information**

Prefix \*First Name M.I. \*Last Name Suffix

\*Email Address (to send order confirmation)

\*Contact Number

\*Mobile Telephone Number \*Last 4 of SSN

**Billing Address**

\*Address Line 1

Address Line 2

\*City \*State \*Zip Code

Select a State

**Shipping Address**  Same as Billing Address

Prefix \*First Name M.I. \*Last Name Suffix

\*Address Line 1

Address Line 2

\*City \*State \*Zip Code

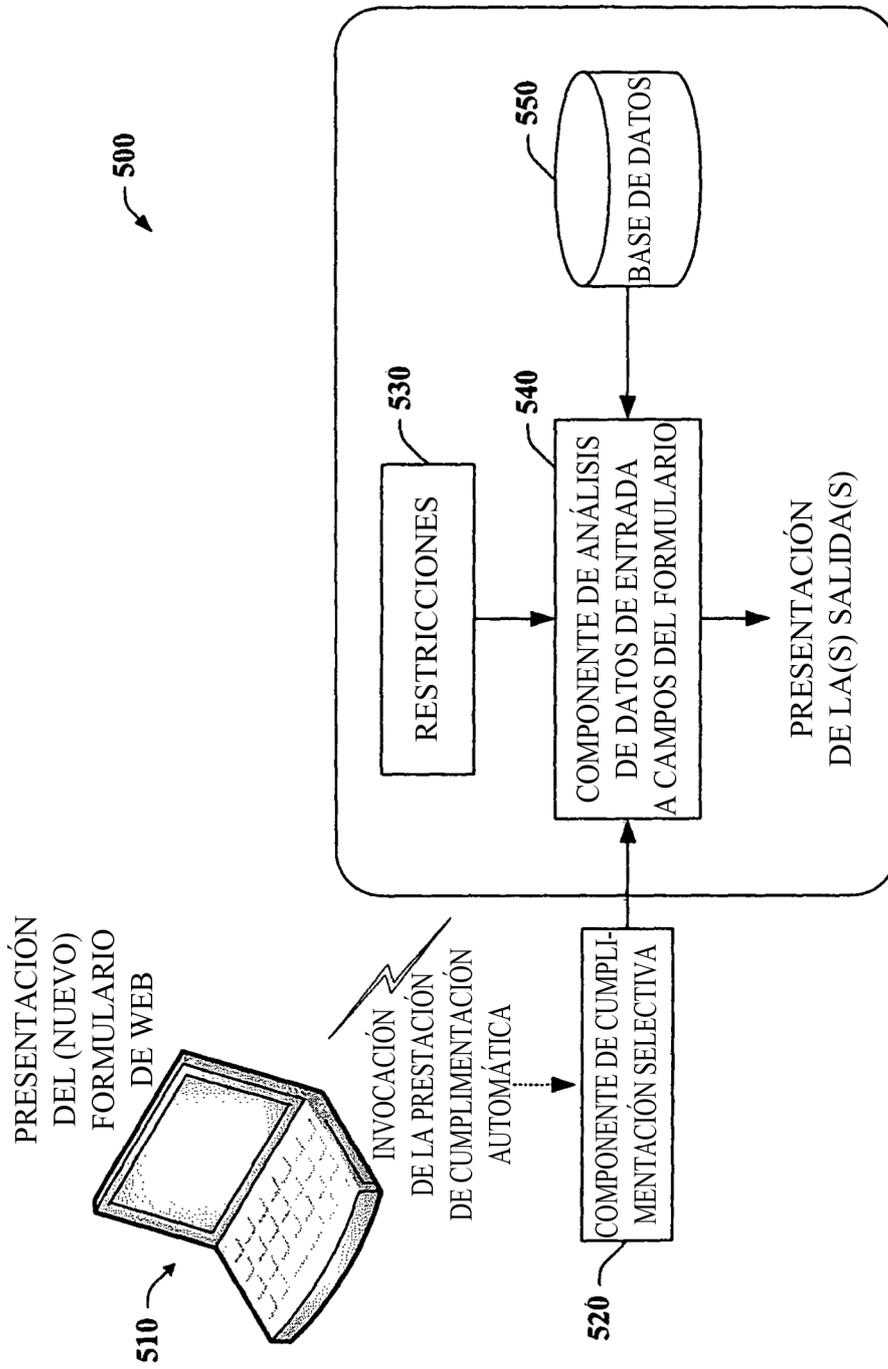
Select a State

**Payment Information**

Your credit card name and billing address must match the billing address entered above.

Done My Computer

FIG. 4



**FIG. 5**

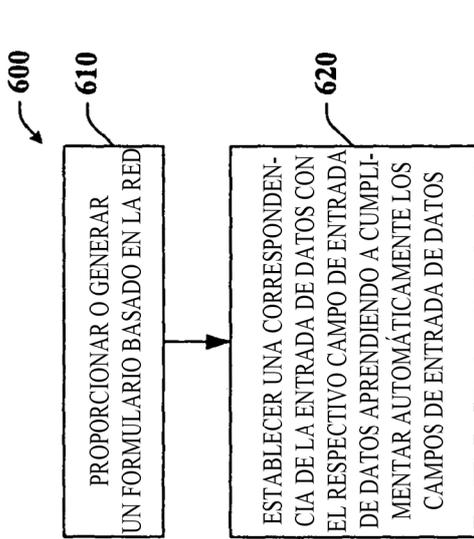


FIG. 6

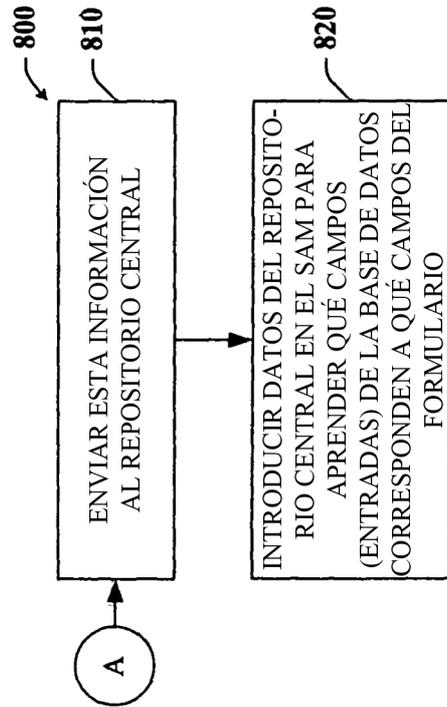


FIG. 8

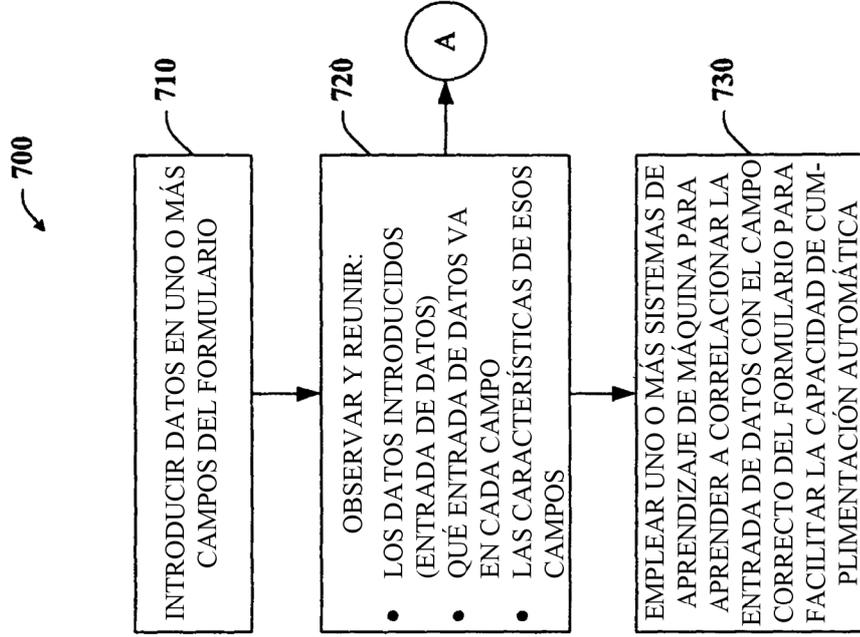


FIG. 7

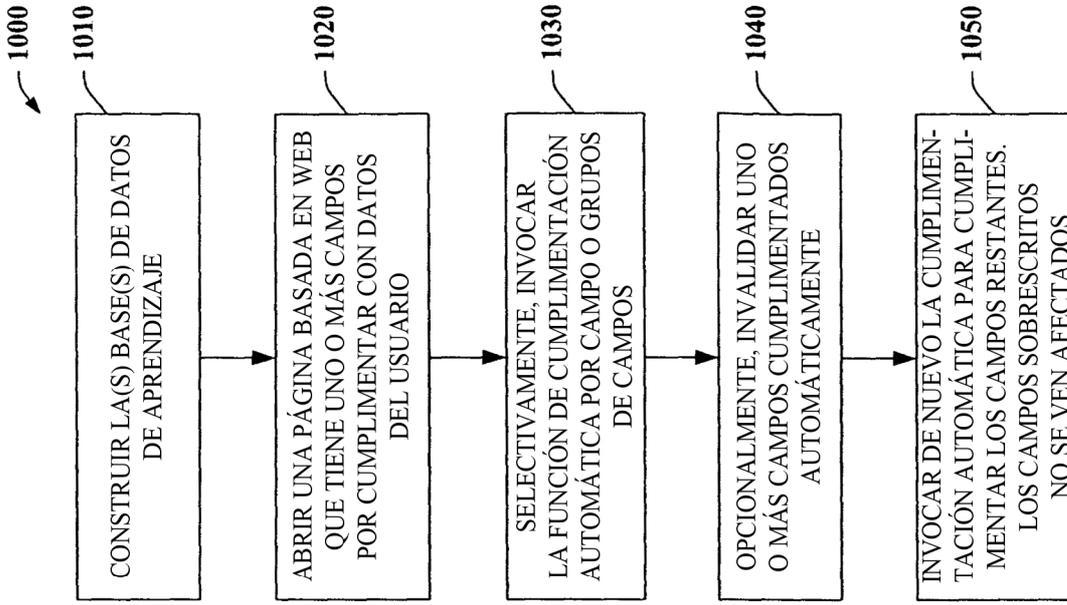


FIG. 10

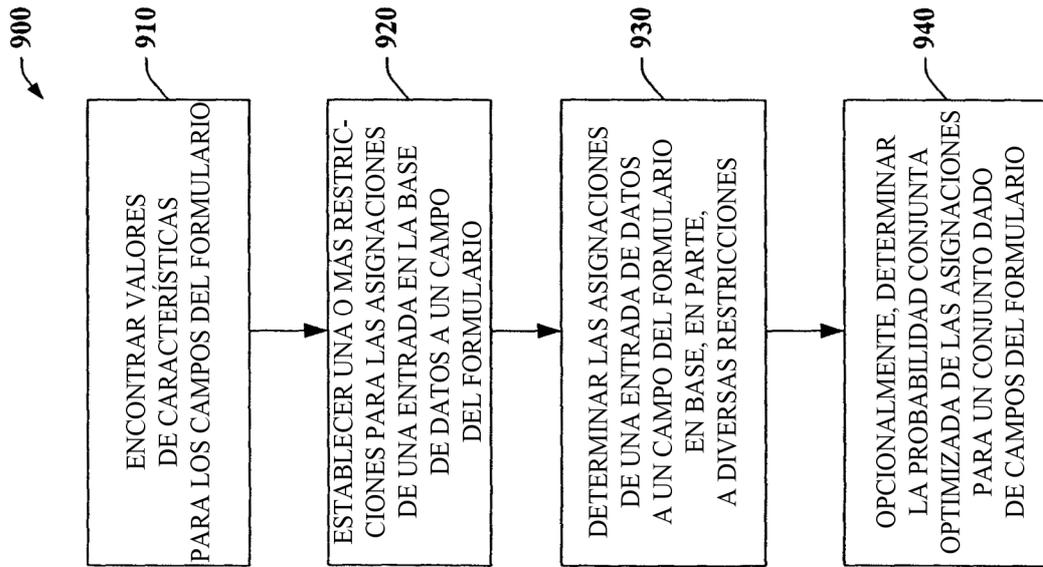


FIG. 9

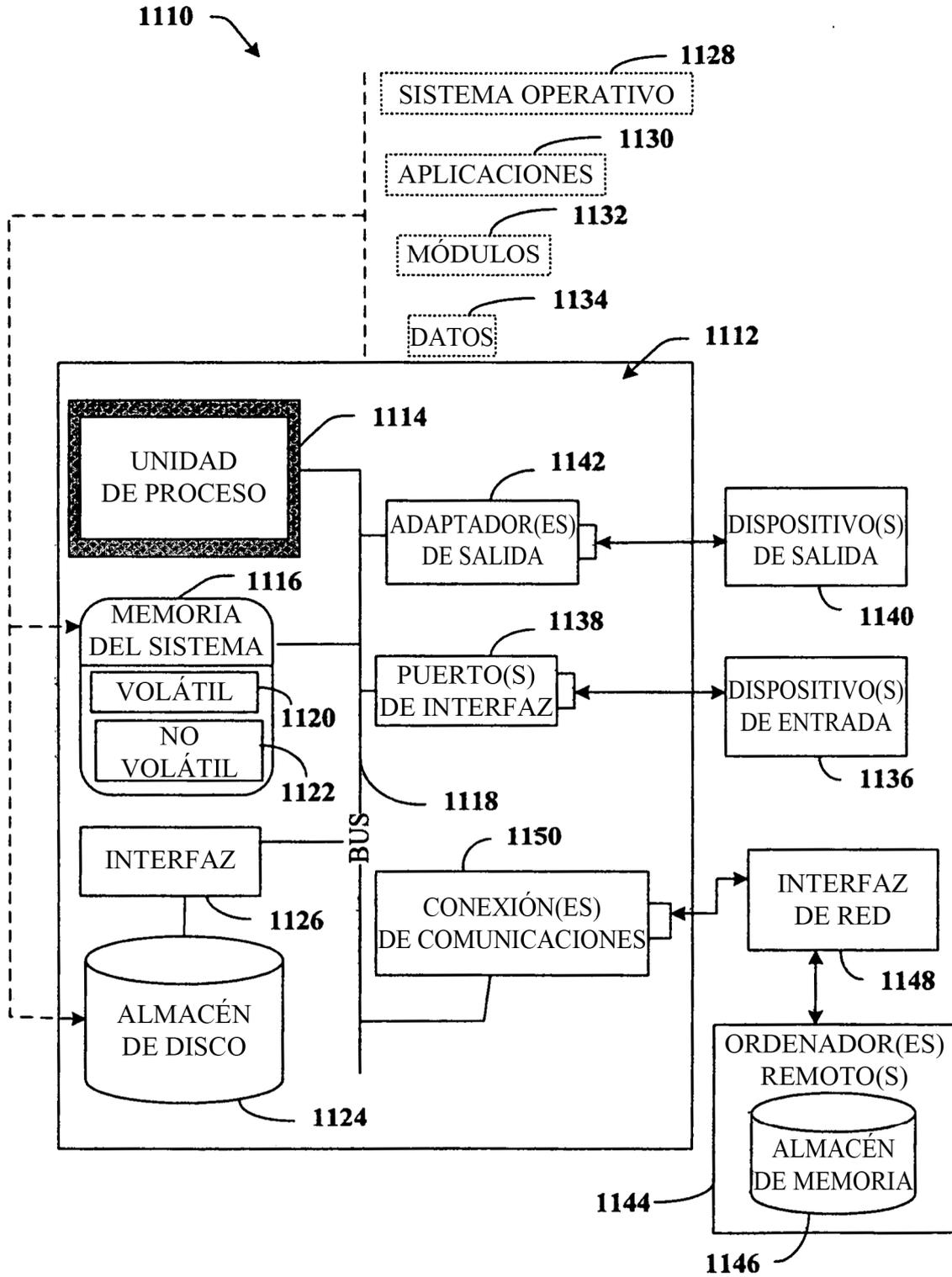


FIG. 11