(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 368 213**

(51) Int. CI.:
G06F 17/24 (2006.01)
G06F 17/27 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **03008805 .8**
(96) Fecha de presentación: **23.04.2003**
(97) Número de publicación de la solicitud: **1361522**
(97) Fecha de publicación de la solicitud: **12.11.2003**

(54) Título: **SISTEMA DE ANOTACIÓN AUTOMÁTICA DE DATOS DE ADIESTRAMIENTO PARA UN SISTEMA DE COMPRENSIÓN DEL LENGUAJE NATURAL.**

(30) Prioridad:
10.05.2002 US 142623

(45) Fecha de publicación de la mención BOPI:
15.11.2011

(45) Fecha de la publicación del folleto de la patente:
15.11.2011

(73) Titular/es:
MICROSOFT CORPORATION
ONE MICROSOFT WAY
REDMOND, WASHINGTON 98052-6399, US

(72) Inventor/es:
Acero, Alejandro;
Wang, Ye-Yi y
Wong, Leon

(74) Agente: **Carpintero López, Mario**

DESCRIPCIÓN

Sistema de anotación automática de datos de adiestramiento para un sistema de comprensión del lenguaje natural

Antecedentes de la invención

La presente invención se refiere a la comprensión del lenguaje natural. Más concretamente, la presente invención se refiere a la anotación de datos de adiestramiento para el adiestramiento de un sistema de comprensión del lenguaje natural.

La comprensión del lenguaje natural es un proceso mediante el cual un usuario de computadora puede suministrar una entrada a una computadora en un lenguaje natural (como por ejemplo mediante una entrada textual o una entrada de voz o por medio de alguna otra interacción con la computadora). La computadora procesa dicha entrada y genera una comprensión de las intenciones que el usuario ha expresado.

Con el fin de adiestrar los sistemas de comprensión del lenguaje natural convencionales, se requiere una gran cantidad de datos de adiestramiento anotados. Sin unos datos de adiestramiento suficientes, los sistemas resultan adiestrados de manera insuficiente y el rendimiento se ve afectado.

Sin embargo, con el fin de generar unos datos de adiestramiento anotados, los sistemas convencionales se basan en anotaciones manuales. Este sistema presenta una pluralidad de inconvenientes significativos. La anotación manual puede ser costosa, retardataria, monótona y propensa al error. Así mismo, pueden incluso resultar difíciles anotaciones de corrección. Si las anotaciones son casi correctas, es bastante difícil detectar errores.

El documento de publicación posterior WO 031096217 A, aplicable solo con arreglo al apartado 3 del Art. 54 del CPE describe una herramienta de desarrollo integrada para construir una aplicación de comprensión del lenguaje natural. La técnica descrita incluye la determinación de la información de la interpretación de NLU a partir de un cuerpo de texto de adiestramiento de NLU utilizando una técnica de procesamiento de múltiples pasadas. La alteración de una pasada puede alterar de forma automática una entrada de una pasada posterior. La información de interpretación de NLU puede especificar una interpretación de al menos parte del cuerpo de texto de adiestramiento de NLU. Los elementos seleccionados de la información de interpretación de NLU pueden ser presentados en un editor gráfico. La información de interpretación de NLU se presenta como un árbol de significados que incluye unos nodos terminales y no terminales. Indicando una probabilidad si puede ser determinada una porción del árbol de significados. Esa porción del árbol de significados puede ser visualmente identificada. El árbol de significados puede ser completado de forma automática de acuerdo con unos datos de anotación predeterminados o con un modelo que especifique unas interpretaciones del texto. Puede llevarse a cabo una determinación acerca de si un único elemento de datos del diccionario de elementos de datos está asociado con una palabra del cuerpo de texto de adiestramiento de NLU. Si es así, el único elemento de datos puede ser asignado a la palabra. Los elementos seleccionados de la información de interpretación de NLU pueden ser representados en forma de sugerencia y una probabilidad que indique si el árbol de significados presentado es una interpretación correcta. La técnica puede, así mismo, incluir la búsqueda de la información de la interpretación de NLU para una estructura específica de árbol de significados. Una intersección de los elementos de datos puede ser identificada y presentada como selecciones para la anotación de una palabra específica para el usuario del cuerpo de texto de adiestramiento de NLU.

El documento: “Aprendizaje para generar una Anotación Semántica para Sentencias Específicas de Dominioo” [“Learning to Generate Semantic Annotation for Domain Specific Sentences”] de Jianming Li, Lei Zhang, Yong Yu, con fecha 21 de octubre de 2001, se refiere al aprendizaje para generar anotaciones semánticas para sentencias específicas de dominio. Las palabras abiertas son anotadas como conceptos en la sentencia, y las palabras cerradas son marcadas para su tratamiento ulterior. En la fase de adiestramiento, todas las palabras abiertas son seleccionadas una por una. Una interfaz de adiestramiento proporciona una lista de posibles conceptos, y el apropiado es elegido entre la lista. Un vector de contexto es generado para cada palabra abierta en la estructura de enlace de las sentencias.

El documento “Comprensión del Lenguaje Natural de Base Estocástica a Través de Tareas y Lenguajes” [“Stochastically-Based Natural Language Understanding Across Tasks and Languages”] de Minker W., con fecha 22 de septiembre de 1997, describe una comprensión del lenguaje natural de base estocástica a través de tareas y lenguajes. Los parámetros del modelo son estimados mediante un procedimiento estocástico que requiere unos cuerpos anotados de forma semántica. Se utiliza una técnica iterativa, semiautomática, para anotar los datos. Manualmente se determinan unos análisis sintácticos para una pluralidad de sentencias. A continuación, se inicia un procedimiento iterativo: utilizando el modelo y anotando la consulta en el subconjunto siguiente. Para la corrección de datos, cada etiqueta semántica de la secuencia tiene que ser verificada. Los conjuntos anotados fueron fusionados y los parámetros del modelo fueron recalculados. Estas etapas fueron reiteradas hasta que el completo conjunto de adiestramiento fue anotado y corregido semánticamente.

Constituye un objetivo de la presente invención proporcionar un sistema de comprensión del lenguaje natural mejorado.

Este objetivo se consigue mediante la materia objeto de las reivindicaciones independientes.

Formas de realización preferentes se definen en las reivindicaciones dependientes.

Sumario de la invención

La presente invención utiliza un sistema de comprensión del lenguaje natural que está siendo actualmente adiestrado para ayudar a la anotación de los datos de adiestramiento para el adiestramiento de ese sistema de comprensión del lenguaje natural. El sistema es inicialmente, de manera opcional, adiestrado utilizando algunos datos iniciales de adiestramiento anotados. A continuación, se proporcionan al sistema unos datos adicionales de adiestramiento no anotados y el sistema propone unas anotaciones a los datos de adiestramiento. Al usuario se le ofrece una oportunidad para confirmar o corregir las anotaciones propuestas, y el sistema es adiestrado con las anotaciones corregidas o verificadas.

En un ejemplo, cuando el usuario interactúa con el sistema, solo se presentan alternativas legales a la anotación propuesta para su selección por parte del usuario.

En otra forma de realización, el sistema de comprensión del lenguaje natural calcula una métrica de confianza, asociada con las anotaciones propuestas. La métrica de confianza puede ser utilizada para marcar unos datos en la anotación propuesta en los cuales el sistema menos confía. Ello atrae la atención del usuario hacia los datos en los que el sistema menos confía.

En otra forma de realización, con el fin de incrementar la velocidad y precisión con las que el sistema propone anotaciones, el usuario puede limitar los tipos de anotaciones propuestos por el sistema de comprensión del lenguaje natural para un subconjunto predeterminado posible de aquellos. Por ejemplo, el usuario puede seleccionar unas categorías o tipos de interpretaciones lingüísticas para su uso por el sistema. Al delimitar de esta forma las posibles anotaciones propuestas por el sistema, se incrementan la velocidad y la precisión del sistema.

En otra forma de realización, el sistema de comprensión del lenguaje natural recibe un conjunto de anotaciones. El sistema, a continuación, examina las anotaciones para determinar si el sistema ha sido ya adiestrado de manera incoherente con las anotaciones. Esto puede ser utilizado para detectar cualquier tipo de incoherencias, incluso estilo de anotación diferentes utilizados por diferentes anotadores (humanos o máquinas). El sistema puede señalizar esto al usuario en un intento por reducir los errores de usuario o las incoherencias de anotación al anotar los datos.

En otro ejemplo, el sistema jerarquiza las anotaciones propuestas en base a la métrica de confianza en orden ascendente (o descendente). Ello identifica para el usuario los datos de adiestramiento en los cuales el sistema menos confía y prioritiza los datos para su procesamiento por el usuario.

El sistema puede, así mismo, escoger las anotaciones propuestas mediante cualquier tipo prediseñado. Ello permite que el usuario procese (por ejemplo, corrija o verifique) todas las anotaciones propuestas de un tipo determinado, de una vez. Ello hace posible una anotación más rápida y estimula un trabajo de anotación más coherente y más preciso.

El sistema actual puede, así mismo, emplear una diversidad de técnicas diferentes para la generación de las anotaciones propuestas. Dichas técnicas pueden ser utilizadas en paralelo, y puede ser empleado un algoritmo de selección para seleccionar la anotación propuesta para su presentación al usuario en base a los resultados de las diferentes técnicas que están siendo utilizadas. Técnicas diferentes presentan intensidades diferentes, y técnicas combinadas pueden, a menudo, producir mejores resultados que cualquier procedimiento de comprensión del lenguaje individual.

De modo similar, la presente invención puede presentar al usuario las diversas porciones de los modelos de comprensión del lenguaje natural que están siendo empleados y que no hayan recibido unos datos de adiestramiento suficientes. Ello hace posible que el usuario identifique diferentes tipos de datos que todavía son necesarios para adiestrar suficientemente los modelos.

Breve descripción de los dibujos

La FIG. 1 es un diagrama de bloques de un entorno en el cual puede ser utilizada la presente invención.

La FIG. 2 es un diagrama de bloques que ilustra un sistema para el adiestramiento de un sistema de comprensión del lenguaje natural de acuerdo con una forma de realización de la presente invención.

La FIG. 3 es un diagrama de flujo que ilustra el funcionamiento global de la presente invención.

La FIG. 4 es un diagrama de bloques más detallado de un sistema para el adiestramiento de un sistema de comprensión de lenguaje natural de acuerdo con una forma de realización de la presente invención.

La FIG. 5 es un diagrama de flujo más detallado que ilustra una operación de la presente invención.

Las FIGS. 6 y 7 son capturas de pantalla que ilustran formas de realización de una interfaz de usuario empleada por la presente invención.

La FIG. 8 es un diagrama de flujo que ilustra el funcionamiento del presente sistema en la adición o supresión de nodos de anotaciones propuestas de acuerdo con una forma de realización de la presente invención.

La FIG. 9 es un diagrama de flujo que ilustra el uso de una diversidad de técnicas de comprensión del lenguaje natural en la proposición de datos anotados de acuerdo con una forma de realización de la presente invención.

Descripción detallada de formas de realización ilustrativas

La presente invención se refiere a la generación de datos de adiestramiento anotados para el adiestramiento de un sistema de comprensión del lenguaje natural. Sin embargo, antes de analizar la presente invención con detalle, se analizará una forma de realización de un entorno en el cual puede ser utilizada la presente invención.

La FIG. 1 ilustra un ejemplo de un entorno 100 de sistema informático en el que la invención puede ser implementada. El entorno 100 de sistema informático es solo un ejemplo de un entorno informático apropiado y no pretende sugerir ninguna limitación en cuanto el alcance del uso o funcionalidad de la invención. Como tampoco debe el entorno informático 100 ser interpretado como dependiente o condicionado a uno cualquiera o a una combinación de componentes ilustrados en el entorno operativo ejemplar 100.

La invención es operativa con otros numerosos entornos o combinaciones de sistemas informáticos de propósito general o de propósito especial. Ejemplos de sistemas informáticos, entornos, y / o configuraciones bien conocidos que pueden ser utilizados para ser empleados con la invención incluyen, pero no se limitan a, las computadoras personales, las computadoras de servidor, los dispositivos de sujeción manual o portátiles, los sistemas de microprocesador, los sistemas basados en microprocesador, los decodificadores, los dispositivos informáticos programables por el consumidor, los PCs de red, las minicomputadoras, las computadoras para gran sistema, los entornos informáticos distribuidos que incluyan cualquiera de los sistemas o dispositivos anteriores, y similares.

La invención puede ser descrita en el contexto general de las instrucciones ejecutables por computadora, como por ejemplo módulos de programa, que sean ejecutados por una computadora. En términos generales, los módulos de programa incluyen rutinas, programas, objetos, componentes, estructuras de datos, etc. que llevan a cabo tareas específicas o implementan tipos de datos abstractos específicos. La invención puede, así mismo, llevarse a la práctica en entornos informáticos distribuidos donde las tareas se lleven a cabo mediante dispositivos de procesamiento remotos que estén enlazados mediante una red de comunicaciones. En un entorno informático distribuido, los módulos de programa pueden estar situados tanto en medios de almacenamiento por computadora locales como remotos incluyendo dispositivos de almacenamiento de memoria.

Con referencia a la FIG. 1, un sistema ejemplar para la implementación de la invención incluye un dispositivo informático de propósito general bajo la forma de una computadora 110. Los componentes de la computadora 110 pueden incluir, pero no se limitan a, una unidad de procesamiento 120, una memoria 130 del sistema y un bus 121 del sistema, que acopla diversos componentes del sistema, incluyendo la memoria del sistema, a la unidad de procesamiento 120. El bus 121 del sistema puede ser cualquiera de los distintos tipos de estructuras de bus que incluyan un bus de memoria o un controlador de memoria, un bus periférico y un bus local que utilice cualquiera de las diversas arquitecturas de bus. A modo de ejemplo, y no de limitación, dichas arquitecturas incluyen el bus de la Arquitectura Estándar del Sector Informático (ISA), el bus de la Arquitectura Microcanal (MCA), el bus del ISA Ampliado (EISA), el bus local de la Asociación de Normalización Electrónica de Vídeo (VESA), y el bus de Interconexión de Componentes Periféricos (PCI), también conocido como bus Mezzanine.

La computadora 110 típicamente incluye una diversidad de medios legibles por computadora. Los medios legibles por computadora pueden ser cualquier medio disponible al que se pueda acceder por la computadora 110 e incluyen tanto medios volátiles como no volátiles, medios extraíbles como no extraíbles. A modo de ejemplo, y no de limitación, los medios legibles por computadora pueden comprender medios de almacenamiento en computadora y medios de comunicación. Medios de almacenamiento en computadora incluyen tanto medios volátiles como no volátiles, extraíbles como no extraíbles, implementados en cualquier procedimiento o técnica de almacenamiento de información, como por ejemplo instrucciones legibles por computadora, estructuras de datos, módulos de programa u otros datos. Los medios de almacenamiento en computadora incluyen, pero no se limitan a , la RAM, la ROM, la EEPROM, la memoria flash u otro sistema técnico de memoria, el CD-ROM, los discos versátiles digitales (DVD) u otro almacenamiento de disco óptico, casetes magnéticas, cinta magnética, almacenamiento de disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio que pueda ser utilizado para almacenar la información deseada y al que se pueda acceder por la computadora 100. Los medios de comunicación típicamente incorporan instrucciones legibles por computadora, estructuras de datos, módulos de programa, u otros datos en una señal de datos modulada, como por ejemplo una portadora WAV u otro mecanismo de transporte que incluye cualquier medio de distribución de información. El término “señal de datos modulada” significa una señal que presenta una o más de sus características fijadas o modificadas para codificar la información en la señal. A modo de ejemplo, y no de limitación, los medios de comunicación incluyen medios cableados, como por ejemplo una red cableada o una conexión cableada punto a punto, y medios inalámbricos como por ejemplo medios acústicos, de

FR, de infrarrojos y otros medios inalámbricos. Combinaciones de cualquiera de los expuestos deben, así mismo, ser incluidos dentro del alcance de los medios legibles por computadora.

La memoria 130 del sistema incluye unos medios de almacenamiento en computadora bajo la forma de una memoria volátil y / o no volátil, como por ejemplo una memoria de solo lectura (ROM) 131 y una memoria de acceso aleatorio (RAM) 132. Un sistema básico 133 de entrada / salida (BIOS), que contiene las rutinas básicas que ayudan a transferir la información de los elementos situados dentro de la computadora 110, como por ejemplo durante la puesta en marcha, están típicamente almacenados en la ROM 131. La RAM 132 típicamente contiene datos y / o módulos de programa que son inmediatamente accseibles a y / o que actualmente están siendo operados por la unidad de procesamiento 120. A modo de ejemplo, y no de limitación, la FIG. 1 ilustra un sistema operativo 134, unos programas de aplicación 135, otros módulos de programa 136 y unos datos de programa 137.

La computadora 110 puede, así mismo, incluir otros medios de almacenamiento por computadora extraíbles / no extraíbles, volátiles / no volátiles. Solo a modo de ejemplo, la FIG. 1 ilustra una unidad de disco duro 141 que lee de o escribe hacia medios magnéticos no extraíbles, no volátiles, una unidad de disco magnético 151 que lee de o escribe hacia un disco magnético extraíble, no volátil 152, y una unidad de disco óptico 155 que lee de o escribe hacia un disco extraíble , no volátil 156, como por ejemplo un CD-ROM u otros medios ópticos. Otros medios de almacenamiento en computadora extraíbles / no extraíbles, volátiles / no volátiles, que pueden ser utilizados en el entorno operativo ejemplar, incluyen, pero no se limitan a, casetes de cintas magnéticas, tarjetas de memoria flash, discos versátiles digitales, cintas de vídeo digital, RAM de estado sólido, ROM de estado sólido, y similares. La unidad de disco duro 141 está típicamente conectada al bus 121 del sistema mediante una interfaz de memoria no extraíble, como por ejemplo la interfaz 140, y la unidad de disco magnético 151 y la unidad de disco óptico 155 están típicamente conectadas al bus 121 del sistema mediante una interfaz de memoria extraíble, como por ejemplo la interfaz 150.

Las unidades y sus medios de almacenamiento por computadora asociados analizados con anterioridad e ilustrados en la FIG. 1, proporcionan el almacenamiento de las instrucciones legibles por computadora, de las estructuras de datos de los módulos de datos y de otros datos con destino a la computadora 110. En la FIG. 1, por ejemplo, la unidad de disco duro 141 se ilustra almacenando el sistema operativo 144, los programas de aplicación 145, otros módulos de programa 146 y los datos de programa 147. Nótese que estos componentes pueden ser o bien los mismos o bien diferentes del sistema operativo 134, de los programas de aplicación 135, de otros módulos de programa 136 y de los datos de programa 137. Al sistema operativo 144, a los programas de aplicación 145, a los otros módulos de programa 146 y a los datos de programa 147 se les otorgan números diferentes en la presente memoria para ilustrar que, como mínimo, son copias diferentes.

Un usuario puede introducir unos comandos y una información en la computadora 110 a través de unos dispositivos de entrada, como por ejemplo un teclado 162, un micrófono 163, y un dispositivo señalador 161, como por ejemplo un ratón, una bola o una tableta táctil. Otros dispositivos de entrada (no mostrados) pueden incluir una palanca de mando, un mando para videojuego, una antena parabólica, un escáner, o similares. Estos y otros dispositivos de entrada están a menudo conectados a la unidad de procesamiento 120 mediante una interfaz 160 de datos introducidos por el usuario que está acoplada al bus del sistema, pero que puede estar conectada mediante otra interfaz y otras estructuras de bus, como por ejemplo un puerto paralelo, un puerto para juegos, o un bus serie universal (USB). Un monitor 191 u otro tipo de dispositivo de presentación está, así mismo, conectado al bus 121 del sistema por medio de una interfaz, como por ejemplo una interfaz de vídeo 190. Además del monitor, las computadoras pueden, así mismo, incluir otros dispositivos de salida periféricos, como por ejemplo unos altavoces 197 y una impresora 196, los cuales pueden estar conectados mediante una interfaz periférica de salida 190.

La computadora 110 puede operar en un entorno de red utilizando conexiones lógicas con una o más computadoras remotas, como por ejemplo una computadora remota 180. La computadora remota 180 puede ser una computadora personal, un dispositivo de sujeción manual, un servidor, un encaminador, un PC de red, un dispositivo homólogo u otro nodo de red común, y típicamente incluye muchos o todos los elementos descritos con anterioridad con respecto a la computadora 110. Las conexiones lógicas mostradas en la FIG. 1 incluyen una red de área local (LAN) 171 y una red de área extensa (WAN) 173 pero, así mismo, pueden incluir otras redes. Dichos entornos de conexión en red son habituales en oficinas, redes de computadoras de ámbito corporativo, intranets e Internet.

Cuando es utilizada en un entorno de conexión a red de una LAN, la computadora 110 es conectada a la LAN 171 a través de una interfaz o adaptador de red 170. Cuando se utiliza en un entorno de conexión a red de una WAN, la computadora 110 típicamente incluye un módem 172 u otro medio para el establecimiento de comunicaciones a lo largo de la WAN 173, como por ejemplo Internet. El módem 172, el cual puede ser interno o externo, puede estar conectado al bus 121 del sistema por medio de la interfaz 160 de introducción de datos por el usuario o por medio de otro mecanismo apropiado. En un entorno de conexión a red, los módulos de programa presentados con relación a la computadora 110 o partes de estos, pueden ser almacenados en el dispositivo de almacenamiento de memoria remoto. A modo de ejemplo, y no de limitación, la FIG. 1 ilustra unos programas de aplicación remota 185 residiendo en una computadora remota 180. Debe apreciarse que las conexiones a red mostradas son ejemplares y que pueden ser utilizados otros medios de establecimiento de un enlace de comunicaciones entre las computadoras.

Debe destacarse que la presente invención puede ser llevada a cabo sobre un sistema informático, como por ejemplo el descrito en la FIG. 1. Sin embargo, la presente invención puede ser llevada a cabo sobre un servidor, una computadora dedicada a la manipulación de mensajes, o sobre un sistema distribuido en el cual diferentes porciones de la presente invención se lleven a cabo sobre partes diferentes del sistema informático distribuido.

La FIG. 2 es un diagrama de bloques que ilustra un sistema 300 para el adiestramiento de un sistema de comprensión del lenguaje natural (NLU) de acuerdo con una disposición. El sistema 300 incluye un sistema 302 de comprensión del lenguaje natural que va a ser adiestrado. El sistema 300 incluye, así mismo, un componente de aprendizaje 304 y una interfaz de corrección o verificación 306 por el usuario. La FIG. 3 es un diagrama de flujo que ilustra el funcionamiento global del sistema 300 mostrado en la FIG. 2.

El sistema NLU 302 es un sistema ilustrativo de comprensión del lenguaje natural que recibe una entrada del lenguaje natural y la procesa de acuerdo con cualquier técnica de procesamiento del lenguaje natural conocida para obtener y generar de salida una indicación en cuanto al significado de la entrada del lenguaje natural. El sistema NLU 302 incluye así mismo, de manera ilustrativa, unos números que deben ser adiestrados con los datos de adiestramiento anotados.

De acuerdo con un ejemplo de la presente invención, el componente de aprendizaje 304 es un componente de adiestramiento que, de manera opcional, recibe los datos de adiestramiento anotados y adiestra los modelos utilizados en el sistema 302 de comprensión del lenguaje natural (NLU). El componente de aprendizaje 304 puede ser cualquier componente de aprendizaje conocido para la modificación o el adiestramiento de los modelos utilizados en el sistema NLU 302, y el actual procedimiento y entorno informático no está confinado a ningún componente de aprendizaje específico 304.

En cualquier caso, el componente de aprendizaje 304 recibe primeramente, de manera opcional, los datos de adiestramiento anotados iniciales 306. Esto se indica mediante el bloque 308 en la FIG. 3. Los datos de adiestramiento anotados iniciales 306, si se utilizan, incluyen los datos iniciales que han sido anotados por el usuario o por otra entidad de conocimiento del dominio y de los modelos utilizados en el sistema NLU 302. El componente de aprendizaje 304 genera así (o adiestra) los modelos del sistema NLU 302. El adiestramiento del sistema NLU en base a los datos de adiestramiento iniciales es opcional y se ilustra mediante el bloque 310 en la FIG. 3.

El sistema NLU 302 es así inicializado y puede generar anotaciones propuestas para los datos no anotados que reciba, aunque la etapa de inicialización no es necesaria. En todo caso, el sistema NLU 302 no está bien adiestrado todavía, y muchas d sus anotaciones serán probablemente incorrectas.

El sistema NLU 302 recibe a continuación unos datos de adiestramiento no anotados (o parcialmente anotados) 312 respecto de los cuales el usuario desea crear anotaciones para el mejor adiestramiento del sistema NLU 302. Debe destacarse que el actual procedimiento y entorno informático puede ser utilizado para generar, así mismo, anotaciones para datos parcialmente anotados, o para datos completos pero incorrectamente anotados. En lo sucesivo, el término “no anotado” se utiliza para incluir todos estos datos para los cuales se desea una anotación adicional. La recepción de datos de adiestramiento no anotados 312 en el sistema NLU 302 se indica mediante el bloque 314 en la FIG. 3.

El sistema NLU 302 a continuación genera las anotaciones propuestas 316 para los datos de adiestramiento no anotados 312. Esto se indica mediante el bloque 318 en la FIG. 3. Las anotaciones propuestas 316 son suministradas a la interfaz de corrección o verificación de usuario 306 para su presentación al usuario. El usuario puede entonces o bien confirmar las anotaciones propuestas 316 o modificarlas. Esto se describe con mayor detalle más adelante en la solicitud, y se indica mediante el bloque 320 en la FIG. 3.

Una vez que el usuario ha corregido o verificado las anotaciones propuestas 316 para obtener las anotaciones corregidas o verificadas 322 , las anotaciones corregidas o verificadas 322 son suministradas al componente de aprendizaje 304. El componente de aprendizaje 304 a continuación adiestra o modifica los modelos utilizados en el sistema NLU 302 en base a las anotaciones corregidas o verificadas 322. Esto se indica mediante el bloque 324 en la FIG. 3.

De esta manera, el sistema NLU 302 ha participado en la generación de los datos de adiestramiento anotados 322 para su uso en el propio adiestramiento. Aunque las anotaciones propuestas 316 que son creadas en base a los datos de adiestramiento no anotados 312 en una fase temprana en el proceso de adiestramiento pueden ser incorrectas, se ha encontrado que es mucho más fácil para el usuario corregir una anotación incorrecta que crear una anotación para datos de adiestramiento no anotados a partir del borrador. De esta manera, la presente invención aumenta la facilidad con la cual pueden ser generados los datos de adiestramiento anotados.

Así mismo, a medida que el proceso continúa y que el sistema NLU 302 resulta mejor adiestrado, las anotaciones propuestas 316 son correctas en un porcentaje más elevado del tiempo, o al menos resultan más correctas. De esta manera, el sistema comienza a obtener grandes beneficios al crear anotaciones propuestas correctas para el propio adiestramiento.

La FIG. 4 es un diagrama de bloques más detallado del sistema de adiestramiento 300 de acuerdo con una posible disposición. La FIG. 4 ilustra el sistema NLU 302 con mayor detalle y, así mismo, ilustra las estructuras de datos asociadas con el sistema 300 también con mayor detalle.

De modo específico, la FIG. 4 muestra que el sistema NLU 302 incluye un componente 350 de comprensión del lenguaje y un modelo de lenguaje 352, el cual podría, por supuesto, ser cualquier otro modelo utilizado con una técnica concreta de comprensión del lenguaje natural. El componente 350 de comprensión del lenguaje incluye de forma ilustrativa uno o más algoritmos de comprensión del lenguaje conocidos utilizados para el análisis sintáctico de los datos de entrada y generar un análisis sintáctico o anotación de salida indicativa del significado o de la intención de los datos de entrada. El componente 350 accede de forma ilustrativa a uno o más modelos 352 en la realización de sus procesos. El modelo de lenguaje 352 se ilustra a modo de ejemplo, aunque pueden ser utilizados también otros modelos de base estadística o gramatical, u otros modelos (como por ejemplo modelos de lenguaje o modelos semánticos).

La FIG. 4 muestra así mismo que la salida del componente 350 de comprensión del lenguaje incluye de manera ilustrativa unas opciones 353 de anotaciones de adiestramiento y una métrica de confianza 354 de anotaciones. Las opciones 353 de anotaciones de adiestramiento incluyen de manera ilustrativa una pluralidad de hipótesis de anotaciones diferentes generadas por el componente 350 para cada entrada de sentencia de adiestramiento o de cada frase de adiestramiento (u otra unidad de entrada) hacia el componente 350. La métrica de confianza 354 de las anotaciones incluyen de manera ilustrativa una indicación en cuanto a la confianza que el componente 350 tiene en las opciones 353 de anotación de datos de adiestramiento asociadas. En una forma de realización, el componente 350 de comprensión del lenguaje es un componente conocido que genera la métrica de confianza 354 como algo supuesto. La métrica de confianza 354 está asociada con cada porción de las opciones 353 de anotación de adiestramiento.

La FIG. 4 muestra, así mismo, un componente 356 de generación de la métrica de cobertura del modelo del lenguaje. El componente 356 está programado de manera ilustrativa para determinar si las partes del modelo 352 han sido suficientemente adiestradas. Al hacerlo, el componente 356 puede identificar de manera ilustrativa el volumen de los datos de adiestramiento que han sido asociados con cada una de las diversas porciones del modelo 352 para determinar si cualquier parte del modelo 352 no ha sido suficientemente cubierta por los datos de adiestramiento. El componente 356 genera de salida la métrica de cobertura 358 del modelo para su acceso por parte de un usuario. De esta manera, si existen porciones del modelo 352 que no han sido adiestradas con las suficientes cantidades de datos de adiestramiento, el usuario puede agrupar datos de adiestramiento adicionales de un tipo determinado con el fin de adiestrar mejor aquellas porciones del modelo 352.

La FIG. 5 es un diagrama de flujo que ilustra con mayor detalle el funcionamiento del sistema. La FIG. 6 ilustra una disposición de una interfaz de usuario 306 empleada y que se analizará en combinación con la FIG. 5. La interfaz de usuario 306 presenta un primer panel 364, un segundo panel 366 y un tercer panal 368. El panel 364 es un árbol de análisis sintáctico representativo del modelo de lenguaje 352 (el cual, tal y como se mencionó con anterioridad, podría ser cualquier otro tipo de modelo). El árbol de análisis sintáctico presenta una pluralidad de nodos 368, 370, 372, 374, 376 y 378. En la forma de realización ejemplar ilustrada en la FIG. 6, cada uno de estos nodos se corresponde con un comando que va a ser reconocido por el modelo de lenguaje 352.

En el ejemplo ilustrado en la FIG. 6, el sistema de comprensión del lenguaje natural que está siendo empleado es uno que facilita el control y la realización de reservas de aerolínea. Por consiguiente, los nodos 368 a 378 son todos representivos de comandos que van a ser reconocidos por la NLU 302 y, por tanto, específicamente modelados por el modelo de lenguaje 352). De esta manera, los comandos mostrados son, por ejemplo, comandos de “Explicar Código”, “Lista de Aeropuertos”, “Mostrar Capacidad”, etc. Cada uno de los nodos incorpora uno o más nodos hijo dependiente de ellos los cuales contienen atributos que definen con mayor amplitud los nodos de comando. Los atributos son, tal y como se ilustra, franjas que son rellenadas con el fin de identificar completamente el nodo de comando que ha sido reconocido o comprendido por el sistema NLU 302. Las franjas pueden, así mismo, a su vez, incorporar sus propias franjas que van a ser rellenadas.

El panel 366 presenta una pluralidad de frases de adiestramiento diferentes (en los datos de adiestramiento 312) las cuales son utilizadas para adiestrar el modelo presentado por el árbol de análisis sintáctico del panel 364. El usuario puede simplemente seleccionar una de estas frases (por ejemplo pinchando en ella con un cursor de ratón) y el sistema 350 aplica la frase de adiestramiento contra el componente 350 de comprensión del lenguaje y del modelo de lenguaje 352 ilustrado en el panel 364. El análisis sintáctico compuesto (o anotación) 316 que el sistema genera se presenta en el panel 368. El campo 380 presenta la frase de adiestramiento seleccionada por el usuario pero, así mismo, permite que el usuario teclee una frase de adiestramiento que no se ha encontrado en la lista del panel 366.

En funcionamiento, tal y como se ilustra en la FIG. 5), el sistema 300 presenta en primer lugar la cobertura de datos de adiestramiento del lenguaje (por ejemplo la métrica de cobertura 358 del modelo) al usuario. Esto se indica mediante el bloque 360 en la FIG. 5. Al generar la métrica de cobertura del modelo, un grupo de reglas del modelo de lenguaje ilustradas en el panel 364, por ejemplo, puede incorporar una pluralidad de secciones o de reglas gramaticales respecto de las cuales se han procesado muy pocos datos de adiestramiento. En ese caso, si la cantidad de datos de adiestramiento no ha alcanzado un umbral preseleccionado o dinámicamente seleccionado, el

sistema realzará de manera ilustrativa mediante un código de color (o con un contraste visual de otro tipo) determinadas porciones del modelo de lenguaje en el panel 364 para indicar la cantidad de datos de adiestramiento que ha sido recogida y procesada para cada sección del modelo. Por supuesto, el contraste visual puede indicar simplemente que la cantidad de datos ha sido suficiente o insuficiente, o pueden descomponerse en niveles adicionales para proporcionar una indicación de grano más fino en cuanto a la cantidad específica de datos de adiestramiento utilizados para adiestrar cada porción del modelo. Este contraste visual puede, así mismo, basarse en el rendimiento del modelo.

Si se han procesado los suficientes datos, y todas las porciones del modelo 352 están adiestradas en la medida suficiente, el proceso de adiestramiento se ha completado. Esto se indica mediante el bloque 362. Sin embargo, si, en el bloque 362, se determina que se necesitan datos de adiestramiento adicionales, entonces los datos de adiestramiento adicionales no anotados 312 son introducidos en el sistema NLU 302. Esto se indica mediante el bloque 363 en la FIG. 5. Los datos de adiestramiento adicionales 312 incluirán de manera ilustrativa una pluralidad de sentencias o frases de adiestramiento u otras unidades lingüísticas.

Cuando el usuario añada datos de adiestramiento 312, tal y como se ilustra en el bloque 363, múltiples frases de adiestramiento o de sentencias de adiestramiento u otras unidades pueden ser aplicadas al sistema NLU 302. El sistema NLU 302 genera, a continuación, unas propuestas de anotación para todos los ejemplos no anotados alimentados a este como datos de adiestramiento 312. Esto se indica mediante el bloque 390. Por supuesto, para cada ejemplo de adiestramiento no anotado, el sistema NLU 302 puede generar una pluralidad de anotaciones de adiestramiento 353 (mostradas en la FIG. 4) junto con la métrica de confianza 354 de las anotaciones asociadas. Si ese es el caso, el sistema NLU 302 elige una de las opciones de adiestramiento 353 como la anotación propuesta 316 que va a ser representada al usuario. Esto se lleva a cabo de manera ilustrativa utilizando la métrica de confianza 354.

En cualquier caso, una vez que las respuestas de anotación para cada uno de los ejemplos de adiestramiento no anotados han sido generados en el bloque 390, el sistema está listo para la interacción por el usuario, bien sea para verificar o bien para corregir las anotaciones propuestas 316. La forma concreta en la cual las anotaciones propuestas 316 son presentadas al usuario depende de la estrategia de procesamiento que puede ser seleccionada por el usuario, tal y como se indica mediante el bloque 392. Si el usuario selecciona el modo manual, el procesamiento simplemente se desplaza hasta el bloque 394. En ese caso, de nuevo con referencia a la FIG. 6, el usuario simplemente selecciona uno de los ejemplos de adiestramiento del panel 366 y el sistema presenta la anotación propuesta 316 para ese ejemplo de adiestramiento en el panel 368.

El sistema puede, así mismo, en una forma de realización, realzar la porción de la anotación representada en el panel 368 que ofrece la métrica de confianza más baja 354. Esto se indica mediante el bloque 396 en la FIG. 5. Puede ser difícil discernir la diferencia entre una anotación que es ligeramente incorrecta y una que es correcta al 100%. El realce de las secciones de confianza baja de la anotación propuesta atrae la atención del usuario hacia las porciones en las que el sistema NLU 302 tiene menos confianza, incrementado de esta manera la probabilidad de que el usuario discierna las anotaciones incorrectas.

Si, en el bloque 392, el usuario desea reducir al mínimo el tiempo de anotación y mejorar la coherencia de la anotación, el usuario selecciona esta circunstancia mediante una pertinente entrada en la NLU 302, y el sistema NLU 302 genera de salida los ejemplos de los datos de adiestramiento en el panel 366 agrupados por su similitud con el ejemplo que se ha actualmente seleccionado. Esto se indica mediante el bloque 398. En otras palabras, se cree que es más fácil para un usuario corregir o verificar anotaciones propuestas y efectuar unas elecciones de anotación más coherentes si el usuario está corrigiendo al mismo tiempo todas las anotaciones propuestas del mismo tipo. De esta manera, en el ejemplo ilustrado en la FIG. 6, si los datos de adiestramiento incluyen 600 ejemplos de adiestramiento para adiestrar el modelo sobre el comando “Mostrar Vuelo” (representado por el nodo 378) el usuario puede desear procesar (bien corregir o verificar) cada uno de estos ejemplos uno después del otro mejor que procesar algunos ejemplos de “Mostrar Vuelo” entremezclados con otros ejemplos de adiestramiento. En este caso, el sistema 302 agrupa las sentencias de adiestramiento de “Mostrar Vuelo” de manera conjunta y las presenta conjuntamente en el panel 366. Por supuesto, existe una diversidad de técnicas diferentes que pueden ser empleadas para agrupar datos de adiestramiento similares, como por ejemplo la agrupación de anotaciones similares y la agrupación de anotaciones con palabras similares, por citar unas pocas. Por consiguiente, cuando el usuario pincha desde un ejemplo al siguiente, el usuario está procesando datos de adiestramiento similares. Una vez que las sentencias de adiestramiento han sido agrupadas y presentadas de esta manera, el procesamiento avanza con respecto al bloque 394 en el cual el usuario selecciona uno de los ejemplos de adiestramiento y el análisis sintáctico, o la anotación para ese ejemplo se muestra en el panel 368.

Si en el bloque 392, el usuario desea potenciar al máximo el beneficio del adiestramiento por ejemplo corregido o verificado por el usuario, el usuario selecciona esta opción mediante una pertinente entrada en el sistema NLU 302. En ese caso, el sistema NLU 302 aporta las sentencias de adiestramiento y las anotaciones propuestas 316 escogidas en base a la métrica de confianza 354 de las anotaciones por orden ascendente. Esto sitúa las sentencias ejemplares en las cuales el sistema tiene menos confianza en la parte superior de la lista. De esta manera, cuando el usuario selecciona y verifica o corrige cada uno de estos ejemplos, el sistema está aprendiendo más de lo que lo haría si estuviera procesando un ejemplo en el que tuviera un grado de confianza elevado. Por supuesto, las

anotaciones y las sentencias de adiestramiento propuestas pueden ser jerarquizadas también en cualquier otro orden, como por ejemplo mediante un valor descendente de la métrica de confianza. La aportación de los ejemplos de adiestramiento jerarquizados mediante la métrica de confianza se indica mediante el bloque 400 en la FIG. 5.

Con independencia de cuál de las tres estrategias de procedimiento seleccione el usuario, en último término al usuario se le presenta en pantalla la información desplegada en la FIG. 6, o una información similar. De esta manera, el usuario debe seleccionar uno de los ejemplos de adiestramiento entre el panel 366 y del árbol de análisis sintáctico (o anotación) para ese ejemplo se aporta de manera ilustrativa en el bloque 368, con sus porciones de confianza más bajas destacadas de manera ilustrativa o de algún modo indicadas al usuario, en base a la métrica de confianza 354 de las anotaciones. Esto se indica mediante el bloque 396 en la FIG. 5.

El usuario a continuación determina si la anotación es correcta, tal y como se indica mediante el bloque 402. Si no es así, el usuario selecciona el segmento de anotación incorrecta en el análisis sintáctico o la anotación presentada en el panel 368 simplemente pinchando en ese segmento, o realzándolo con el cursor. La selección del segmento de anotación incorrecta se indica mediante el bloque 404 en la FIG. 5.

En el ejemplo mostrado en la FIG. 6, puede apreciarse que el usuario ha realzado el nodo superior de la anotación propuesta (el nodo de “Explicar Código”). Una vez que ese segmento ha sido seleccionado, o realzado, el sistema 302 presenta, por ejemplo en la cuadrícula inferior 410, todas las elecciones de anotación legales (a partir de las opciones 353 de anotación de los datos de adiestramiento) disponibles para el segmento realzado de la anotación. Estas opciones de anotación 353 pueden ser presentadas en la cuadrícula inferior 410 por orden de confianza en base a la métrica de confianza 354 de la anotación, o también en cualquier otro orden deseado. Esto se indica mediante el bloque 412 en la FIG. 5.

El término “elecciones de anotaciones legales” significa aquellas selecciones que no violan las limitaciones del modelo o modelos 352 que son utilizados por el sistema 302. Por ejemplo, para el procesamiento de una entrada en lengua inglesa, el modelo o los modelos 352 pueden efectivamente encontrar limitaciones que indiquen que cada sentencia debe tener un verbo, o que cada frase preposicional debe empezar con una preposición. Dichas limitaciones pueden ser también semánticas. Por ejemplo, las limitaciones pueden permitir una ciudad en el comando “Lista de Aeropuerto” pero no en el comando “Mostrar Capacidad”. Puede utilizarse también cualquier diversidad distinta de limitaciones. Cuando el usuario ha seleccionado una porción de la anotación en el panel 368 que es incorrecta, el sistema 302 no genera todos los posibles análisis sintácticos o anotaciones para ese segmento de los datos de adiestramiento. Por el contrario el sistema 302 solo genera y presenta esas porciones o anotaciones, para ese segmento de los datos de adiestramiento, lo que se traducirá en un análisis sintáctico legal de toda la sentencia de adiestramiento. Si una anotación concreta no pudiera traducirse en un análisis sintáctico global (uno que no viole las limitaciones de los modelos que se están utilizando) entonces el sistema 302 no presenta ese posible análisis sintáctico o anotación como una opción para el usuario en la cuadrícula inferior 410.

Una vez que las alternativas se muestran en la cuadrícula inferior 410, el usuario selecciona la correcta simplemente realzándola y pinchando en ella. Esto se indica mediante el bloque 414 en la FIG. 5. El procesamiento a continuación vuelve a bloque 402 donde se determina que la anotación es ahora correcta.

La anotación corregida o verificada 322 es a continuación guardada y presentada al componente de aprendizaje 304. Esto se indica mediante el bloque 416 en la FIG. 5. El componente de aprendizaje 304 es, de manera ilustrativa, un algoritmo de aprendizaje conocido que modifica el modelo en base a una pieza nuevamente introducida de los datos de adiestramiento (como por ejemplo las anotaciones corregidas o verificadas 322). Los parámetros actualizados del modelo de lenguaje se ilustran mediante el bloque 420 en la FIG. 4, y el proceso de generación de esos parámetros se indica mediante el bloque 422 en la FIG. 5.

El sistema 302 puede, así mismo, verificar las incoherencias entre los datos de adiestramiento previamente anotados. Por ejemplo, el sistema NLU 302 aprende, puede aprender que los datos de adiestramiento anterior o actualmente anotados fueron incorrectamente anotados. Básicamente, ello verifica si el sistema predice de forma correcta las anotaciones que el usuario eligió respecto de ejemplos de adiestramiento pasados. Los errores de la predicción pueden sugerir la contradicción del conjunto del adiestramiento.

La determinación sobre si verificar estas contradicciones puede ser seleccionada por el usuario y se indica mediante el bloque 424. Si el componente de aprendizaje 304 está preparado para verificar contradicciones, el sistema 302 es controlado para de nuevo generar de salida unas anotaciones propuestas para los datos de adiestramiento que han sido ya anotados por el usuario. El componente de aprendizaje 304 compara los datos de la anotación guardada (la anotación que fue verificada o corregida por el usuario y guardada) con las anotaciones generadas de manera automática. El componente de aprendizaje 304 a continuación busca incoherencias en las dos anotaciones, tal y como se indica mediante el bloque 430. Si no hay incoherencias, entonces ello significa que las anotaciones corregidas o verificadas por el usuario no se consideran erróneas por el sistema y el procesamiento simplemente vuelve el bloque 390 donde las propuestas de anotación son generadas para el siguiente ejemplo no anotado seleccionado por el usuario.

Sin embargo, si, en el bloque 430, se encuentran incoherencias, esto significa que el sistema 302 ha sido ya adiestrado en un volumen suficiente de datos de adiestramiento que produciría una anotación incoherente con la anteriormente verificada o corregida por el usuario de forma que el sistema presenta un grado bastante alto de confianza de que la entrada del usuario fue incorrecta. De esta manera, el procesamiento de nuevo vuelve al bloque 396 donde la anotación corregida o verificada del usuario es de nuevo presentada por el usuario en el panel 368, de nuevo con las porciones de confianza baja realzadas para dirigir la atención del usuario hacia la porción de la anotación que el sistema 302 ha considerado probablemente errónea. Esto proporciona al usuario otra oportunidad para verificar la anotación para asegurar que es correcta tal y como se ilustra en el bloque 402.

Cuando las anotaciones han sido finalmente verificadas o corregidas, el usuario puede simplemente pinchar en el botón “Aprender este Análisis Sintáctico” (u otro accionador similar) dispuesto sobre la UI 306 y el modelo de lenguaje es actualizado por el componente de aprendizaje 304.

Debe, así mismo, ser destacado que también se contempla otra característica distintiva. Incluso si se generan y se presentan solo anotaciones legales al usuario durante la corrección, ello puede tardar una cantidad considerable de tiempo. Por tanto, se dispone un mecanismo mediante el cual el usuario puede limitar el análisis del lenguaje natural del ejemplo de entrada a subconjuntos específicos de los análisis posibles. Dichos límites pueden, por ejemplo, referirse a la limitación del análisis a una sola categoría lingüística o a una determinada porción del modelo. En el ejemplo ilustrado en la FIG. 6, si el usuario está procesando los comandos “Mostrar Capacidad”, por ejemplo, el usuario puede simplemente realzar esa porción del modelo antes de seleccionar una sentencia de adiestramiento siguiente. Esto se indica mediante los bloques 460 y 462 en la FIG. 5. De esta manera, en las etapas en las que el sistema NLU 302 está generando las anotaciones propuestas, limitará su análisis y propuestas solo a aquéllas anotaciones que se incluyan en el nodo seleccionado del modelo. En otras palabras, el sistema NLU 302 intentará solo establecer una correspondencia de la sentencia de adiestramiento de entrada con los nodos con arreglo al nodo de comando realzado. Ello puede reducir de manera considerable la cantidad de tiempo de procesamiento y mejorar la precisión a la hora de generar las anotaciones propuestas.

En las FIGS. 7 y 8, ilustran otra característica distintiva adicional de acuerdo con una disposición adicional. Tal y como se analizó con anterioridad con respecto a la FIG. 6, una vez que el usuario selecciona una sentencia de adiestramiento del panel 366, esa sentencia o frase de adiestramiento se aplica por referencia al modelo de lenguaje (u otro modelo) del sistema NLU 302 (o ha sido ya aplicado) y el sistema 302 genera una anotación propuesta que se presenta en el panel 368. Si esa anotación propuesta es incorrecta, el usuario puede realzar el segmento incorrecto de la anotación y el sistema presentará todas las alternativas legales. Sin embargo, puede suceder que una porción de la propuesta de anotación presentada en el panel 368 puede ser incorrecta no precisamente porque un nodo esté incorrectamente identificado, sino porque falta un nodo y debe ser añadido, o porque existen muchos nodos y uno debe ser borrado o dos deben ser combinados.

Si un nodo debe ser borrado, el usuario simplemente lo realza y a continuación selecciona borrar en la cuadrícula inferior 410. Sin embargo, si deben efectuarse cambios adicionales en la estructura del nodo, el usuario puede seleccionar la opción “añadir hijo” en la cuadrícula inferior 410. En ese caso, al usuario se le ofrece una imagen similar a la mostrada en la FIG. 7.

La FIG. 7 ofrece un primer campo 500 y un segundo campo 502. El primer campo 500 presenta una porción de la sentencia de adiestramiento o de la frase de adiestramiento, y realza la porción de la frase de adiestramiento que ya no resulta cubierta por la propuesta de anotación presentada en el panel 368, pero que debería estarlo. Debe apreciarse que las anotaciones completas no necesitan cubrir todas y cada una de las palabras de una sentencia de adiestramiento. Por ejemplo, la palabra “Por Favor” que precede a un comando puede no ser anotada. La actual característica distintiva de la invención simplemente se aplica a porciones no cubiertas por la anotación, pero que pueden ser necesarias para obtener una anotación completa.

En el ejemplo mostrado en la FIG. 7, en la porción de la sentencia de adiestramiento que se presenta es “Seattle a Boston”. La FIG. 7 ilustra así mismo que el término “Seattle” está cubierto por la anotación actualmente seleccionada en el panel 368 porque “Seattle” aparece en letras grises. Los términos “a Boston” aparecen en negrita indicando que todavía no están cubiertos por el análisis sintáctico (o anotación) actualmente representado en el panel 368.

La cuadrícula 502 ilustra todas las opciones de anotación legales disponibles para los términos “a Boston”. El usuario puede simplemente seleccionar uno de aquellos realzándolo y accionando el botón “ok”. Sin embargo, el usuario puede, así mismo, realzar una u otra o ambas palabras (“a Boston”) de la cuadrícula 500, y el sistema 302 genera todas las posibles opciones de anotación legales para las palabras realzadas, y presenta estas opciones en la cuadrícula 302. De esta manera, si el usuario selecciona “a”, la cuadrícula 502 relacionará todas las posibles anotaciones legales para “a”. Si el usuario selecciona “Boston”, la cuadrícula 502 relaciona todas las anotaciones legales para “Boston”. Si el usuario selecciona “a Boston”, la cuadrícula 502 relaciona todas las anotaciones legales para “a Boston”. De esta manera, el usuario puede diversificar la porción de la sentencia de adiestramiento representada en la cuadrícula 500 (la cual no está actualmente cubierta por la anotación propuesta) en cualquier número deseado de nodos, simplemente realzando cualquier número de porciones de la sentencia de adiestramiento y seleccionar la adecuada de las opciones de anotación legal representadas en la cuadrícula 502.

De manera específica, tal y como se muestra en la FIG. 8, supóngase que el sistema 300 ha presentado un análisis sintético o anotación propuesta en el panel 368 para una sentencia de adiestramiento seleccionada. Esto se indica mediante el bloque 504. Supóngase entonces que el usuario ha borrado un nodo hijo incorrecto, tal y como se indica mediante el bloque 506. El usuario, a continuación, selecciona la opción “Añadir Hijo” en la cuadrícula inferior 410’ tal y como se indica mediante el bloque 508. Esto genera una presentación similar a la mostrada en la FIG. 7, en la que el sistema presenta la porción de los datos de adiestramiento no cubiertos todavía por el análisis sintético (dado que parte de la anotación o del análisis sintáctico propuesto ha sido borrado por el usuario). Esto se indica mediante el bloque 510.

El sistema, a continuación, presenta las alternativas legales para una porción seleccionada de los datos de adiestramiento no cubiertos, tal y como se indica mediante el bloque 512. Si el usuario selecciona una de las alternativas, entonces la anotación presentada en el panel 368 es corregida en base a la selección del usuario. Esto se indica mediante los bloques 514 y 516 y se determina si la anotación actual está completa. Si no, el procesamiento vuelve al bloque 510. Si es así, el procedimiento se completa con respecto a esta sentencia de adiestramiento. Esto se indica mediante el bloque 518.

Si, en el bloque 514, el usuario no seleccionó una de las alternativas del campo 502, entonces se determina si el usuario ha seleccionado (o realzado) una porción de los datos de adiestramiento no cubiertos del campo 500. Si no, el sistema simplemente espera a que el usuario, o bien seleccione una parte de los datos no cubiertos presentados en el campo 500, o bien seleccione un análisis sintáctico correcto del campo 502. Esto se indica mediante el bloque 520. Sin embargo, si el usuario ha realzado una porción de los datos de adiestramiento no cubiertos en el campo 500, entonces el procesamiento retorna al bloque 512 y el sistema presenta las alternativas legales de los datos de adiestramiento no cubiertos seleccionados, para que el usuario pueda seleccionar la anotación apropiada.

De acuerdo con otro ejemplo adicional, se conoce una diversidad de técnicas diferentes para generar anotaciones de sentencias (o cualquier otra unidad del lenguaje natural, como por ejemplo una palabra, un grupo de palabras, una(s) frase(s) o una sentencia o grupo de sentencias). Por ejemplo, son conocidos sistemas de clasificación tanto estadísticos como basados en la gramática para generar anotaciones a partir de entradas del lenguaje natural. De acuerdo con una disposición, una pluralidad de técnicas diferentes es utilizada para generar anotaciones para las mismas sentencias de adiestramiento (u otras unidades del lenguaje natural). El sistema 302 incluye por tanto, en el componente 350 de comprensión del lenguaje, una diversidad de algoritmos diferentes para generar anotaciones propuestas. Por supuesto, el sistema 302 incluye así mismo, de manera ilustrativa, los correspondientes modelos asociados con aquellos diferentes algoritmos. La FIG. 9 es un diagrama de flujo que ilustra la forma en que estas técnicas y diferentes algoritmos y modelos pueden ser utilizadas de acuerdo con una disposición.

El usuario, en primer término, indica al sistema 302 (mediante un accionador de la interfaz de usuario u otra técnica de entrada) cuál de las técnicas de generación de anotaciones desea emplear el usuario (todas, algunas o sola una). Esto se indica mediante el bloque 600. Las técnicas pueden ser elegidas comprobando el rendimiento de cada una con respecto a las sentencias de anotación personal o cualquier otra forma de determinar cuáles son las más eficaces. El sistema 300, a continuación, adiestra los modelos asociados con cada una de aquellas técnicas acerca de los datos de adiestramiento iniciales anotados para inicializar el sistema. Esto se indica mediante el bloque 602. Los modelos adiestrados son, a continuación, utilizados para proponer anotaciones para los datos de adiestramiento no anotados de manera similar a la descrita con anterioridad, estribando la diferencia en que las anotaciones son generadas utilizando una pluralidad de técnicas diferentes. Esto se indica mediante el bloque 604.

Los resultados de las diferentes técnicas son, a continuación, combinados para elegir una anotación propuesta para su presentación al usuario. Esto se indica mediante el bloque 606. Una amplia variedad de algoritmos de combinación puede ser utilizada para escoger la anotación apropiada. Por ejemplo, puede ser empleado un algoritmo de votación para escoger la anotación propuesta sobre la cual esté de acuerdo la mayoría de las técnicas de generación de anotaciones. Por supuesto, pueden ser utilizadas otras combinaciones similares e incluso más brillantes para escoger una anotación propuesta de aquellas generadas por la técnica de generación de anotaciones.

Una vez que la anotación concreta ha sido escogida, como la anotación propuesta, se presenta por medio de la interfaz de usuario. Esto se indica mediante el bloque 608.

De esta manera, puede apreciarse que pueden ser utilizados muchas formas de realización y ejemplos diferentes con el fin de facilitar la anotación oportuna, eficiente y de poco coste de los datos de adiestramiento con el fin de adiestrar un sistema de comprensión del lenguaje natural. Simplemente la utilización del propio sistema NLU para generar las propuestas de anotación reduce drásticamente la cantidad de tiempo y de trabajo manual requerido para anotar los datos de adiestramiento. Aun cuando el sistema a menudo producirá inicialmente errores, es menos difícil corregir una anotación propuesta que lo es el crear una anotación a partir de un borrador.

Mediante la presentación de solo alternativas legales en el curso de la corrección, el sistema promueve una edición de anotación más eficiente. De modo similar, utilizando una métrica de confianza para centrar la atención del usuario sobre porciones de las anotaciones propuestas respecto de las cuales el sistema tiene una confianza menor, reduce

los errores de anotación y reduce la cantidad de tiempo requerido para verificar una propuesta de anotación correcta.

Así mismo, mediante la interpretación de una interfaz de usuario que permite que un usuario limite los procedimientos de comprensión del lenguaje natural a unos subconjuntos del modelo mejora, así mismo, el rendimiento. Si el usuario está anotando un conjunto de datos pertenecientes a una sola característica lingüística, el usuario puede limitar el análisis del lenguaje natural a esa categoría y acelerar el procesamiento, y mejorar la precisión de las propuestas de anotación.

Posibles disposiciones pueden, así mismo, contribuir a discernir los errores de anotación del usuario mediante la aplicación del algoritmo de comprensión del lenguaje a los datos de adiestramiento anotados (confirmados o corregidos por el usuario) y realzar los supuestos en los que el sistema está en desacuerdo con la anotación, o simplemente presentando la anotación con una métrica de confianza baja realzada. El sistema puede, así mismo, ser configurado para prioritizar el adiestramiento sobre datos de confianza baja. En una disposición, esos datos de adiestramiento son presentados al usuario para su procesamiento en primer término.

En otro ejemplo, los datos de adiestramiento similares son agrupados en conjunto utilizando propuestas de anotación generadas de forma automática o cualquier otra técnica para caracterizar la similitud lingüística. Esto hace más fácil para el usuario anotar los datos de adiestramiento, porque el usuario está anotando al mismo tiempo ejemplos de adiestramiento similares. Esto, así mismo, permite que el usuario anote de manera más coherente con menos errores. Así mismo, los patrones de los datos de adiestramiento pueden ser más fáciles de identificar cuando se agrupen ejemplos de adiestramiento idénticos.

Una posible disposición proporciona, así mismo, la combinación de múltiples algoritmos de comprensión del lenguaje natural (o técnicas de generación de propuestas de anotación) para obtener unos resultados más precisos. Estas técnicas pueden ser utilizadas en paralelo para mejorar la cantidad del soporte de anotación suministrada al usuario.

Así mismo, dado que, en general, es importante obtener datos de adiestramiento que cubran todas las porciones del modelo del lenguaje (o de otro modelo que está siendo utilizado), una disposición presenta una representación del modelo del lenguaje y realza o contrasta visualmente porciones del modelo en base a la cantidad de datos de adiestramiento que han sido utilizados en el adiestramiento de esas porciones. Esto puede guiar al usuario en sus esfuerzos de recogida de datos de adiestramiento para indicar qué porciones del modelo necesitan más los datos de adiestramiento.

Aunque la presente invención ha sido descrita con referencia a formas de realización concretas, los expertos en la materia advertirán que pueden llevarse a cabo modificaciones en cuanto a forma y detalle sin apartarse del alcance de la presente invención.

REIVINDICACIONES

1.- Un procedimiento para la generación de datos de adiestramiento anotados para adiestrar un sistema de comprensión del lenguaje natural, NLU, que incorpora uno o más modelos, comprendiendo el procedimiento:

la generación (390) de una anotación propuesta con el sistema NLU para cada unidad de datos de adiestramiento no anotados;

la presentación de las anotaciones propuestas para la verificación o la corrección de usuario para obtener una anotación confirmada por un usuario;

el adiestramiento del sistema NLU con la anotación confirmada por un usuario;

la presentación de una indicación de un volumen de datos de adiestramiento utilizados para adiestrar una pluralidad de porciones diferentes de los uno o más modelos del sistema de comprensión del lenguaje natural;

en el que la presentación de la anotación propuesta para la verificación o corrección del usuario comprende:

la recepción de una entrada de usuario indicativa de una porción identificada por un usuario de la anotación propuesta; y

la presentación de una pluralidad de anotaciones alternativas propuestas para la porción identificada por el usuario;

en el que los uno o más modelos imponen unas restricciones del modelo y en el que la presentación de una o más anotaciones alternativas propuestas comprende la presentación de una anotación alternativa propuesta para la porción identificada por el usuario, solo si la anotación alternativa propuesta puede conducir a una anotación global para la unidad que sea compatible con las restricciones de modelo;

en el que la anotación propuesta incluye unos nodos padres e hijos y en el que la presentación de una pluralidad de anotaciones alternativas propuestas incluye la presentación de una entrada de nodo de borrado accionable por el usuario, la cual, cuando es accionada, borra un nodo hijo, y una entrada de nodo de adición accionable por el usuario, la cual, cuando es accionada, añade un nodo hijo, y la presentación de la pluralidad de anotaciones de alternativas propuestas en respuesta al borrado por parte del usuario de un nodo hijo asociado con la porción identificada por el usuario de la anotación propuesta;

en el que la presentación de una pluralidad de anotaciones alternativas propuestas comprende la presentación de una porción de la unidad no cubierta por la anotación propuesta; y la presentación de anotaciones alternativas propuestas para la porción no cubierta por la anotación propuesta;

en el que el usuario es habilitado para seleccionar un segmento de la porción de la unidad no cubierta por la anotación propuesta y en el que la presentación de las anotaciones alternativas propuestas, comprende la presentación de una o más anotaciones alternativas propuestas para el segmento seleccionado por el usuario; y

en el que el usuario es habilitado para seleccionar una de las anotaciones alternativas propuestas entre la pluralidad de anotaciones alternativas propuestas, y la anotación alternativa propuesta seleccionada por el usuario es incorporada en los datos de adiestramiento anotados.

2.- El procedimiento de la reivindicación 1, y que así mismo comprende: la inicialización de uno o más modelos del sistema NLU.

3.- El procedimiento de la reivindicación 1, en el que la presentación de una indicación de un volumen de datos de adiestramiento comprende:

la presentación de una representación de los uno o más modelos; y

la contrastación visual de porciones de los uno o más modelos que han sido adiestrados con un volumen de umbral de datos de adiestramiento.

4.- El procedimiento de la reivindicación 3, en el que el volumen de umbral de los datos de adiestramiento es dinámico, en base a uno o más criterios de rendimiento para los uno o más modelos.

5.- El procedimiento de la reivindicación 1, en el que la presentación de la anotacion propuesta para su verificación o corrección comprende:

la generación de una métrica de confianza para la anotación propuesta; y

la contrastación visual de una porción de la anotación propuesta presentada en base a la métrica de confianza.

6.- El procedimiento de la reivindicación 5, en el que la contrastación visual comprende:

la contrastación visual de la porción de la anotación presentada que tiene una métrica de confianza que está por debajo de un nivel de umbral.

7.- El procedimiento de la reivindicación 1 y que así mismo comprende:

antes de la generación de una anotación propuesta, la recepción de una indicación de usuario limitativa; y

la limitación del procesamiento de comprensión del lenguaje natural utilizado para generar la anotación propuesta en base a la indicación de usuario limitativa.

8.- El procedimiento de la reivindicación 7, en el que la limitación del procesamiento de comprensión del lenguaje natural, comprende:

la limitación del procesamiento de comprensión del lenguaje natural a la utilización de solo porciones identificadas por el usuario de los uno o más modelos.

9.- El procedimiento de la reivindicación 1 y que comprende así mismo: la identificación de las incoherencias entre la anotación confirmada por el usuario y las anotaciones anteriores.

10.- El procedimiento de la reivindicación 9 y comprende así mismo:

si se identifica una incoherencia, la presentación de la anotación confirmada por el usuario que contrasta visualmente con las porciones incoherentes de la anotación confirmada por el usuario.

11.- El procedimiento de la reivindicación 1, en el que la presentación de las anotaciones propuestas comprende:

la generación de una métrica de enseñanza para cada anotación propuesta;

la presentación de las anotaciones propuestas en un orden en base a la métrica de confianza.

12.- El procedimiento de la reivindicación 1, en el que la presentación de las anotaciones propuestas comprende:

la elección de las anotaciones propuestas en base al tipo de anotación.

13.- El procedimiento de la reivindicación 12, en el que la presentación de las anotaciones propuestas comprende:

la presentación de tipos similares de anotaciones estrechamente próximas entre sí.

14.- El procedimiento de la reivindicación 12, en el que la presentación de las anotaciones propuestas comprende:

la provisión de una entrada accionable por el usuario, la cual, cuando es accionada, permite que el usuario corrija o verifique tipos similares de anotaciones de manera secuencial.

15.- El procedimiento de la reivindicación 1, en el que la generación de anotaciones propuestas por el sistema NLU comprende:

la generación de una pluralidad de anotaciones para cada unidad que utiliza una pluralidad de sistemas NLU diferentes.

16.- El procedimiento de la reivindicación 15, en el que la generación de las anotaciones propuestas comprende así mismo:

la elección de una de las anotaciones propuestas para cada unidad que va a ser presentada.

17.- Un entorno informático que comprende un procesador, estando el entorno informático configurado para ejecutar una interfaz de usuario adaptada para llevar a cabo el procedimiento de acuerdo con lo definido en una de las reivindicaciones 1 a 16.

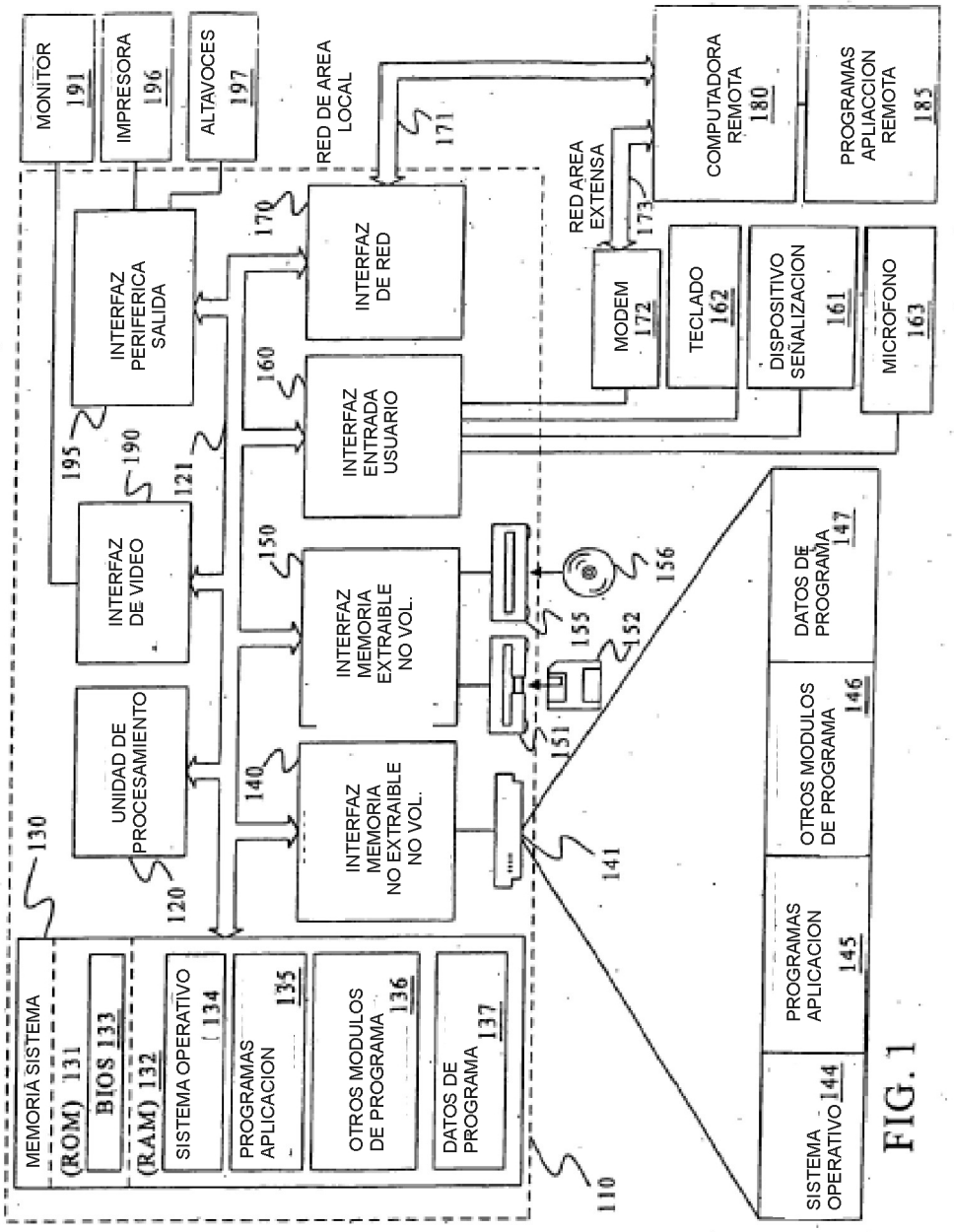

MONITOR 191
IMPRESORA 196
ALTAVOCES 197
RED DE AREA LOCAL
171
INTERFAZ PERIFERICA SALIDA
INTERFAZ DE RED
170
195
190
160
INTERFAZ ENTRADA USUARIO
INTERFAZ DE VIDEO
121
150
INTERFAZ MEMORIA EXTRAIBLE NO VOL.
UNIDAD DE PROCESAMIENTO
140
INTERFAZ MEMORIA NO EXTRAIBLE NO VOL.
120
130
MEMORIA SISTEMA
(ROM) 131
BIOS 133
(RAM) 132
SISTEMA OPERATIVO 134
PROGRAMAS APLICACION 135
OTROS MODULOS DE PROGRAMA 136
DATOS DE PROGRAMA 137
110
141
151
155
152
156
RED AREA EXTENSA
173
MODEM 172
TECLADO 162
DISPOSITIVO SEÑALIZACION 161
MICROFONO 163
COMPUTADORA REMOTA 180
PROGRAMAS APLIACCION REMOTA 185
DATOS DE PROGRAMA 147
OTROS MODULOS DE PROGRAMA 146
PROGRAMAS APLICACION 145
SISTEMA OPERATIVO 144

FIG. 1

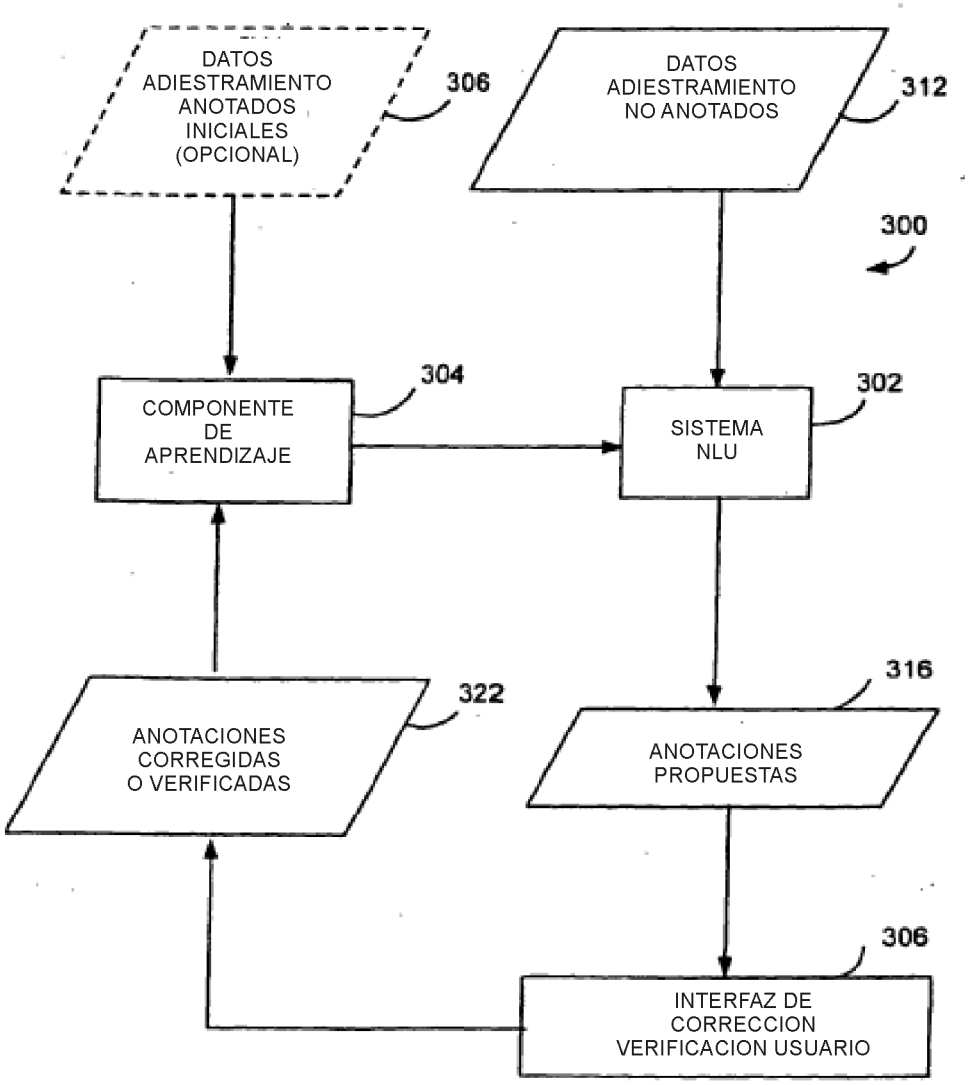

DATOS ADIESTRAMIENTO ANOTADOS INICIALES (OPCIONAL)
306
DATOS ADIESTRAMIENTO NO ANOTADOS
312
300
COMPONENTE DE APRENDIZAJE
304
SISTEMA NLU
302
ANOTACIONES CORREGIDAS O VERIFICADAS
322
ANOTACIONES PROPUESTAS
316
INTERFAZ DE CORRECCION VERIFICACION USUARIO
306

FIG. 2

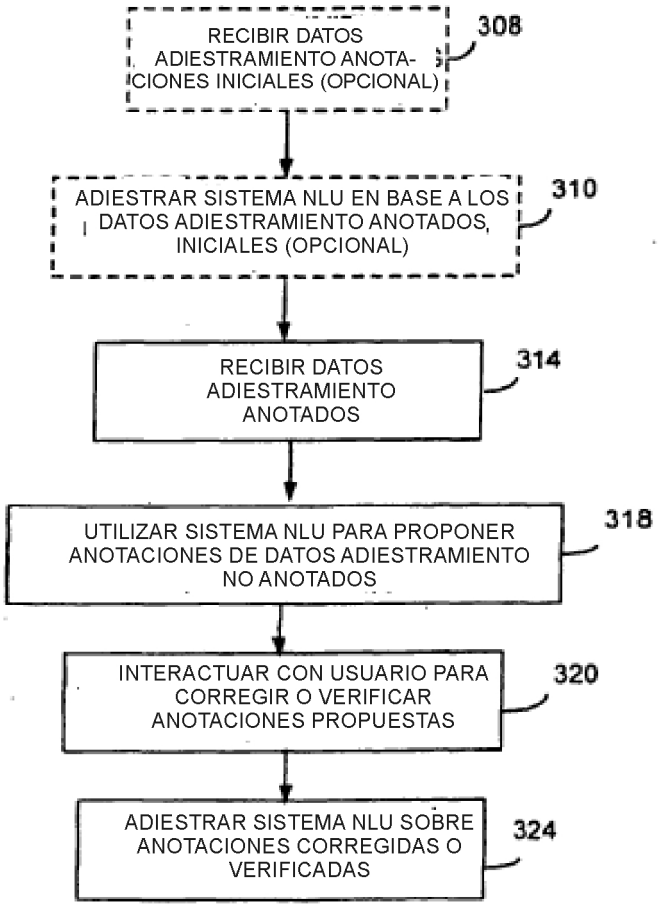

RECIBIR DATOS ADIESTRAMIENTO ANOTA-CIONES INICIALES (OPCIONAL)
308
ADIESTRAR SISTEMA NLU EN BASE A LOS DATOS ADIESTRAMIENTO ANOTADOS, INICIALES (OPCIONAL)
310
RECIBIR DATOS ADIESTRAMIENTO ANOTADOS
314
UTILIZAR SISTEMA NLU PARA PROPONER ANOTACIONES DE DATOS ADIESTRAMIENTO NO ANOTADOS
318
INTERACTUAR CON USUARIO PARA CORREGIR O VERIFICAR ANOTACIONES PROPUESTAS
320
ADIESTRAR SISTEMA NLU SOBRE ANOTACIONES CORREGIDAS O VERIFICADAS
324

FIG. 3

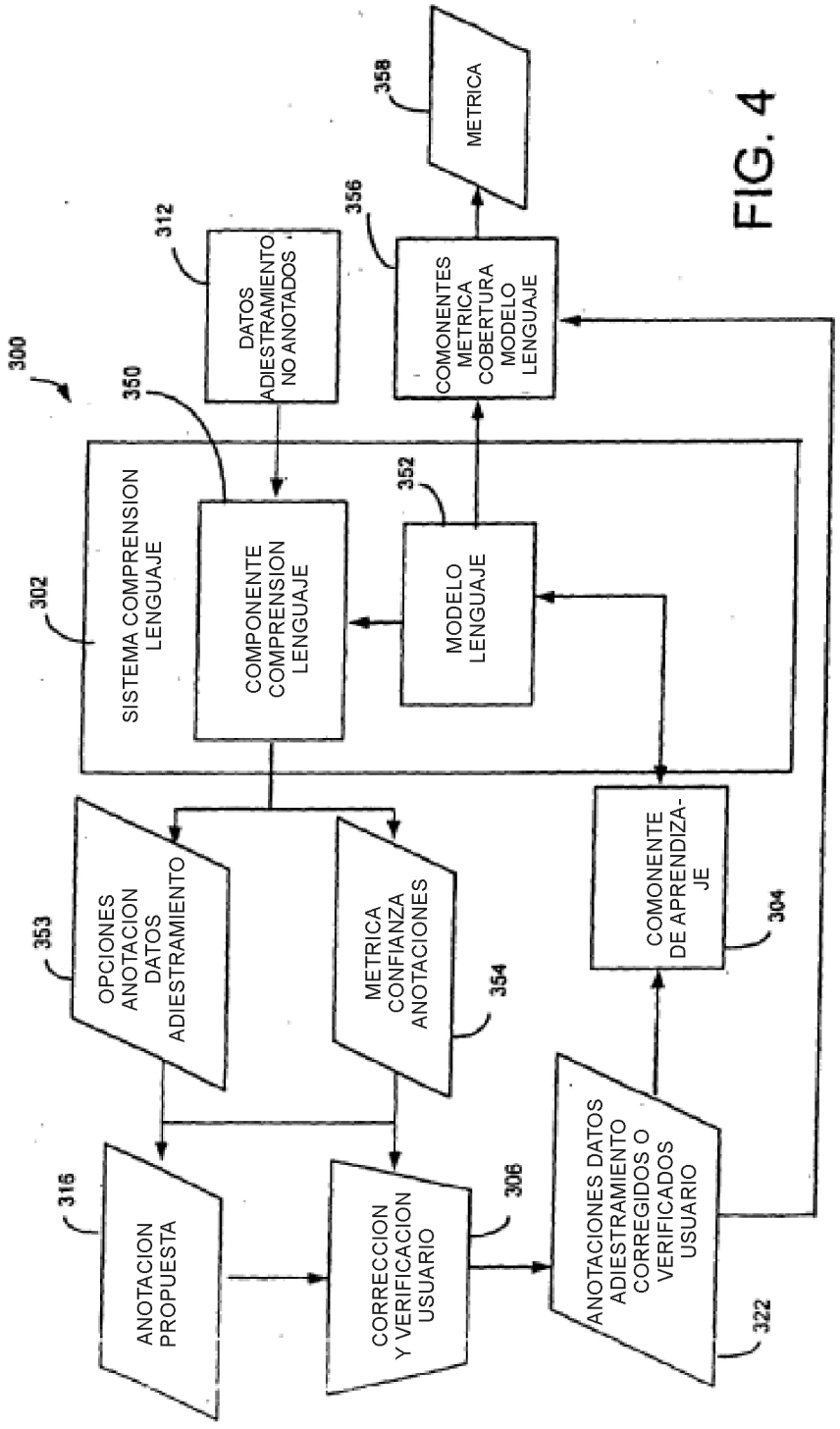

300
302
SISTEMA COMPRENSION LENGUAJE
350
COMPONENTE COMPRENSION LENGUAJE
352
MODELO LENGUAJE
312
DATOS ADIESTRAMIENTO NO ANOTADOS
356
COMONENTES METRICA COBERTURA MODELO LENGUAJE
358
METRICA
353
OPCIONES ANOTACION DATOS ADIESTRAMIENTO
354
METRICA CONFIANZA ANOTACIONES
304
COMONENTE DE APRENDIZA-JE
316
ANOTACION PROPUESTA
306
CORRECCION Y VERIFICACION USUARIO
322
ANOTACIONES DATOS ADIESTRAMIENTO CORREGIDOS O VERIFICADOS USUARIO

FIG. 4

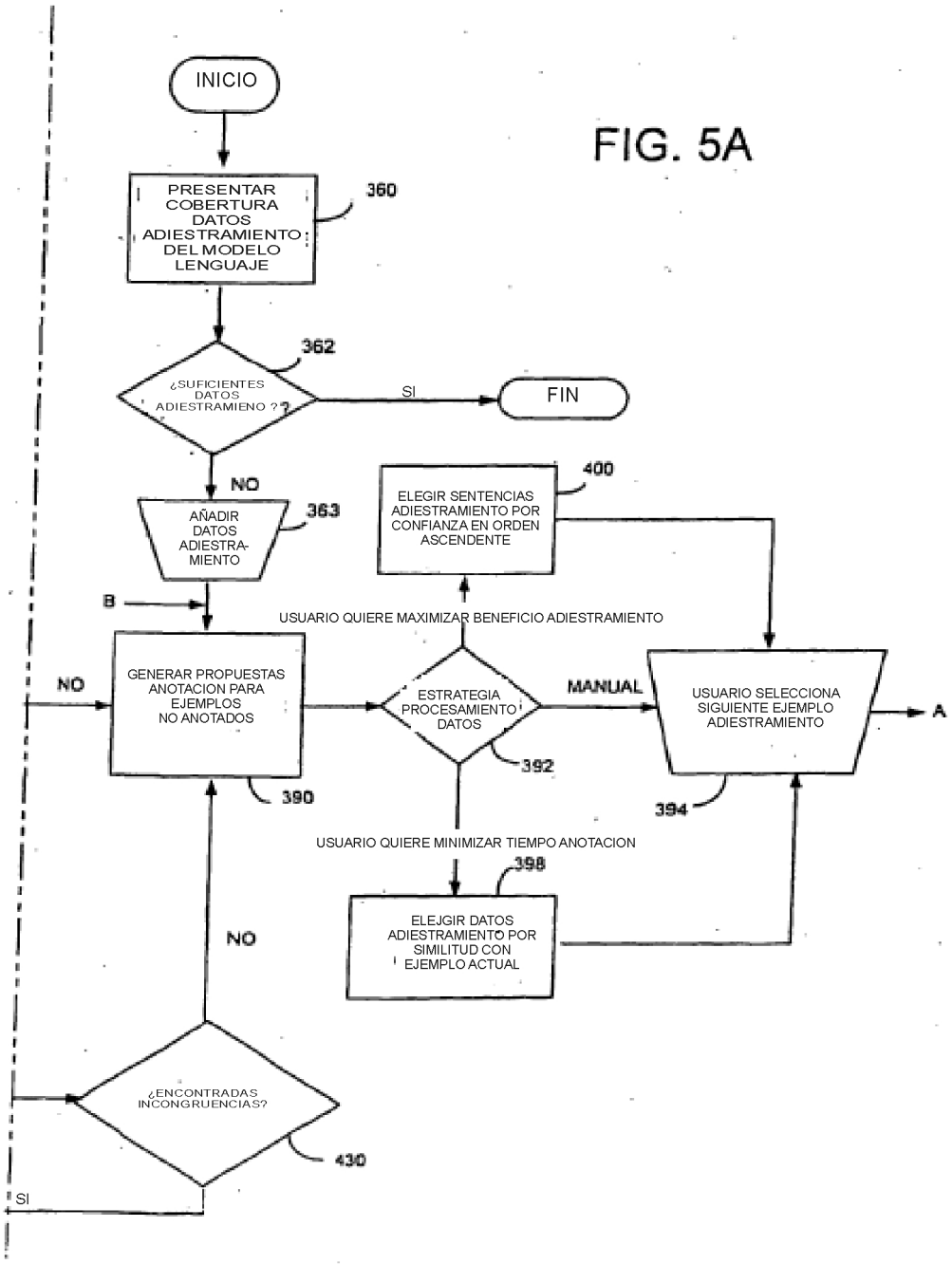
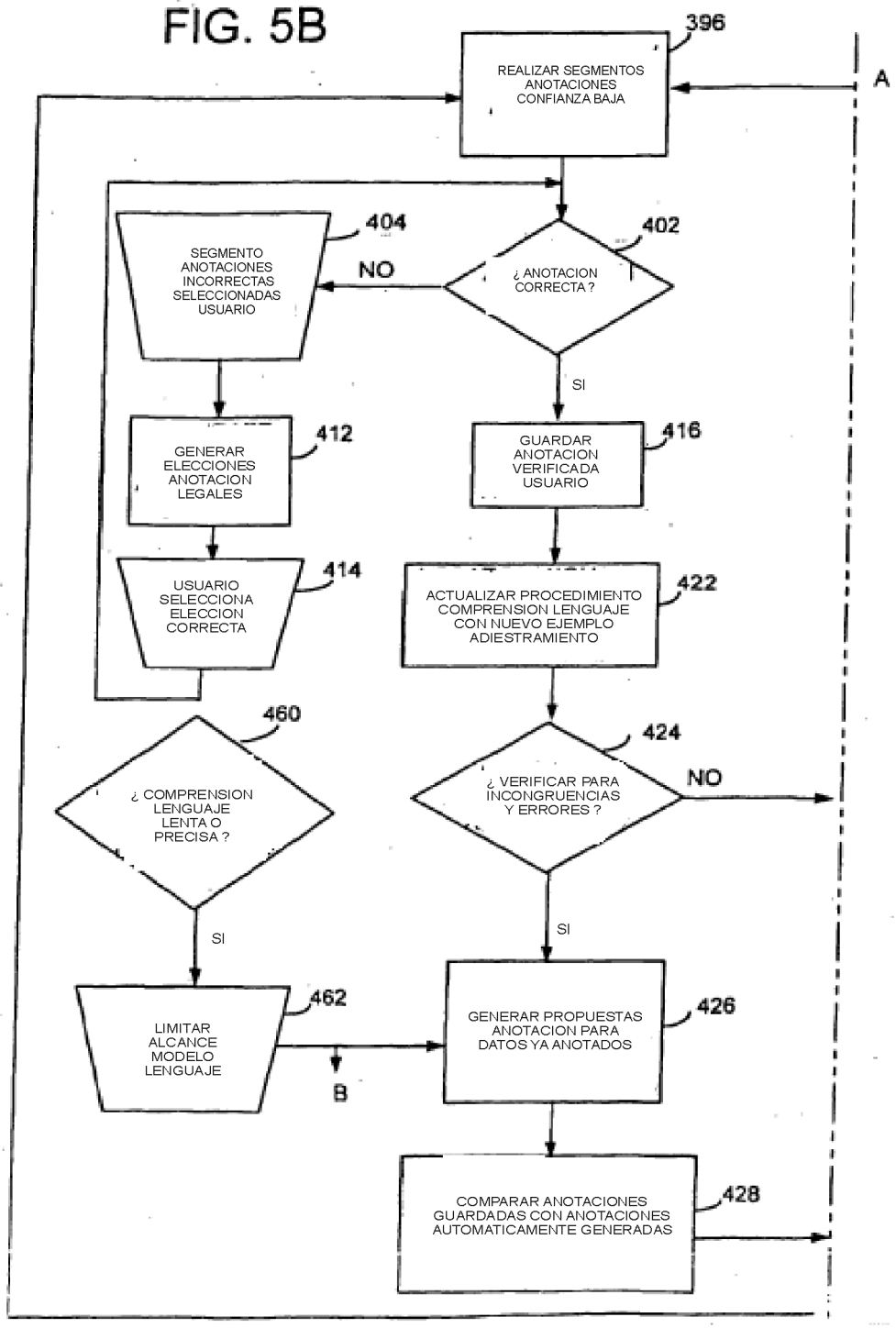

INICIO
PRESENTAR COBERTURA DATOS ADIESTRAMIENTO DEL MODELO LENGUAJE
360
362
¿SUFICIENTES DATOS ADIESTRAMIENO ??
SI
FIN
NO
AÑADIR DATOS ADIESTRA-MIENTO
363
400
ELEGIR SENTENCIAS ADIESTRAMIENTO POR CONFIANZA EN ORDEN ASCENDENTE
B
USUARIO QUIERE MAXIMIZAR BENEFICIO ADIESTRAMIENTO
NO
GENERAR PROPUESTAS ANOTACION PARA EJEMPLOS NO ANOTADOS
390
ESTRATEGIA PROCESAMIENTO DATOS
392
MANUAL
USUARIO SELECCIONA SIGUIENTE EJEMPLO ADIESTRAMIENTO
394
A
USUARIO QUIERE MINIMIZAR TIEMPO ANOTACION
398
ELEJGIR DATOS ADIESTRAMIENTO POR SIMILITUD CON EJEMPLO ACTUAL
NO
¿ENCONTRADAS INCONGRUENCIAS?
430
SI

FIG. 5A

FIG. 5B


396
REALIZAR SEGMENTOS ANOTACIONES CONFIANZA BAJA
A
404
SEGMENTO ANOTACIONES INCORRECTAS SELECCIONADAS USUARIO
NO
402
¿ ANOTACION CORRECTA ?
SI
412
GENERAR ELECCIONES ANOTACION LEGALES
416
GUARDAR ANOTACION VERIFICADA USUARIO
414
USUARIO SELECCIONA ELECCION CORRECTA
422
ACTUALIZAR PROCEDIMIENTO COMPRENSION LENGUAJE CON NUEVO EJEMPLO ADIESTRAMIENTO
460
¿ COMPRENSION LENGUAJE LENTA O PRECISA ?
424
¿ VERIFICAR PARA INCONGRUENCIAS Y ERRORES ?
NO
SI
SI
462
LIMITAR ALCANCE MODELO LENGUAJE
B
426
GENERAR PROPUESTAS ANOTACION PARA DATOS YA ANOTADOS
428
COMPARAR ANOTACIONES GUARDADAS CON ANOTACIONES AUTOMATICAMENTE GENERADAS

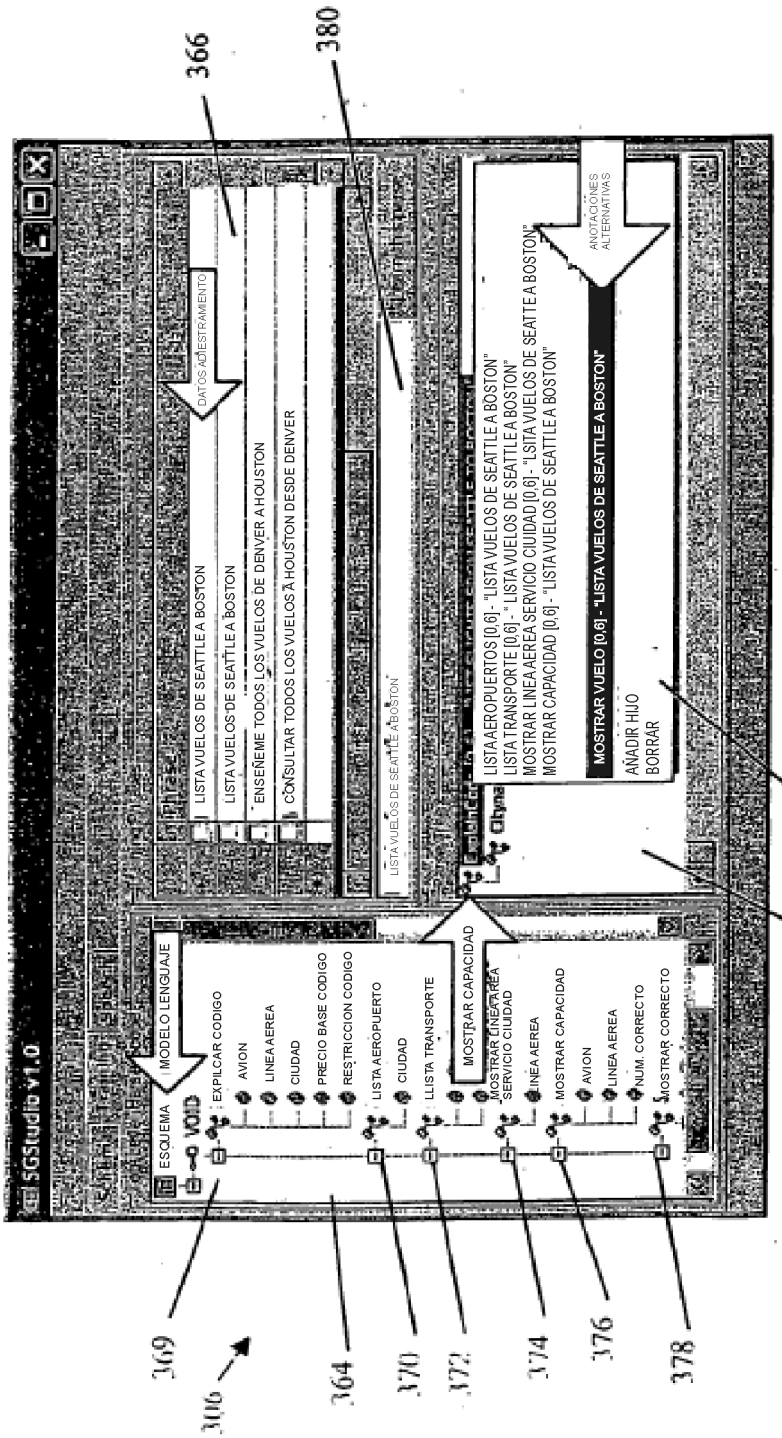

366
380
SGStudio v1.0
ESQUEMA
MODELO LENGUAJE
VOID
EXPLICAR CODIGO
AVION
LINEA AEREA
CIUDAD
PRECIO BASE CODIGO
RESTRICCION CODIGO
LISTA AEROPUERTO
CIUDAD
LLISTA TRANSPORTE
MOSTRAR CAPACIDAD
MOSTRAR LINEA AEREA SERVICIO CIUDAD
LINEA AEREA
MOSTRAR CAPACIDAD
AVION
LINEA AEREA
NUM. CORRECTO
MOSTRAR CORRECTO
DATOS ADIESTRAMIENTO
LISTA VUELOS DE SEATTLE A BOSTON
LISTA VUELOS DE SEATTLE A BOSTON
ENSÉÑEME TODOS LOS VUELOS DE DENVER A HOUSTON
CONSULTAR TODOS LOS VUELOS A HOUSTON DESDE DENVER
LISTA VUELOS DE SEATTLE A BOSTON
LISTA AEROPUERTOS [0,6] - "LISTA VUELOS DE SEATTLE A BOSTON"
LISTA TRANSPORTE [0,6] - " LISTA VUELOS DE SEATTLE A BOSTON"
MOSTRAR LINEA AEREA SERVICIO CIUDAD [0,6] - "LSITA VUELOS DE SEATTLE A BOSTON"
MOSTRAR CAPACIDAD [0,6] - "LISTA VUELOS DE SEATTLE A BOSTON"
MOSTRAR VUELO [0,6] - "LISTA VUELOS DE SEATTLE A BOSTON"
AÑADIR HIJO
BORRAR
ANOTACIONES ALTERNATIVAS
410
368
369
306
364
370
372
374
376
378

FIG. 6

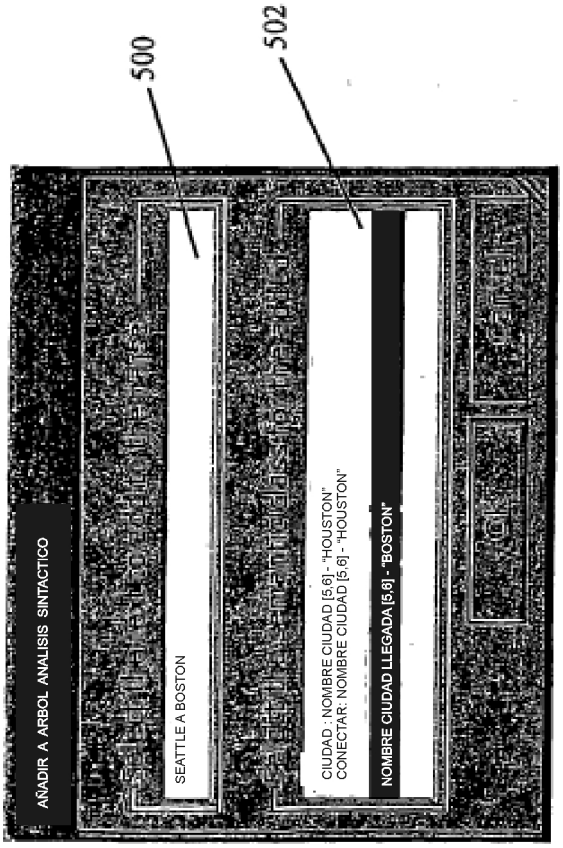
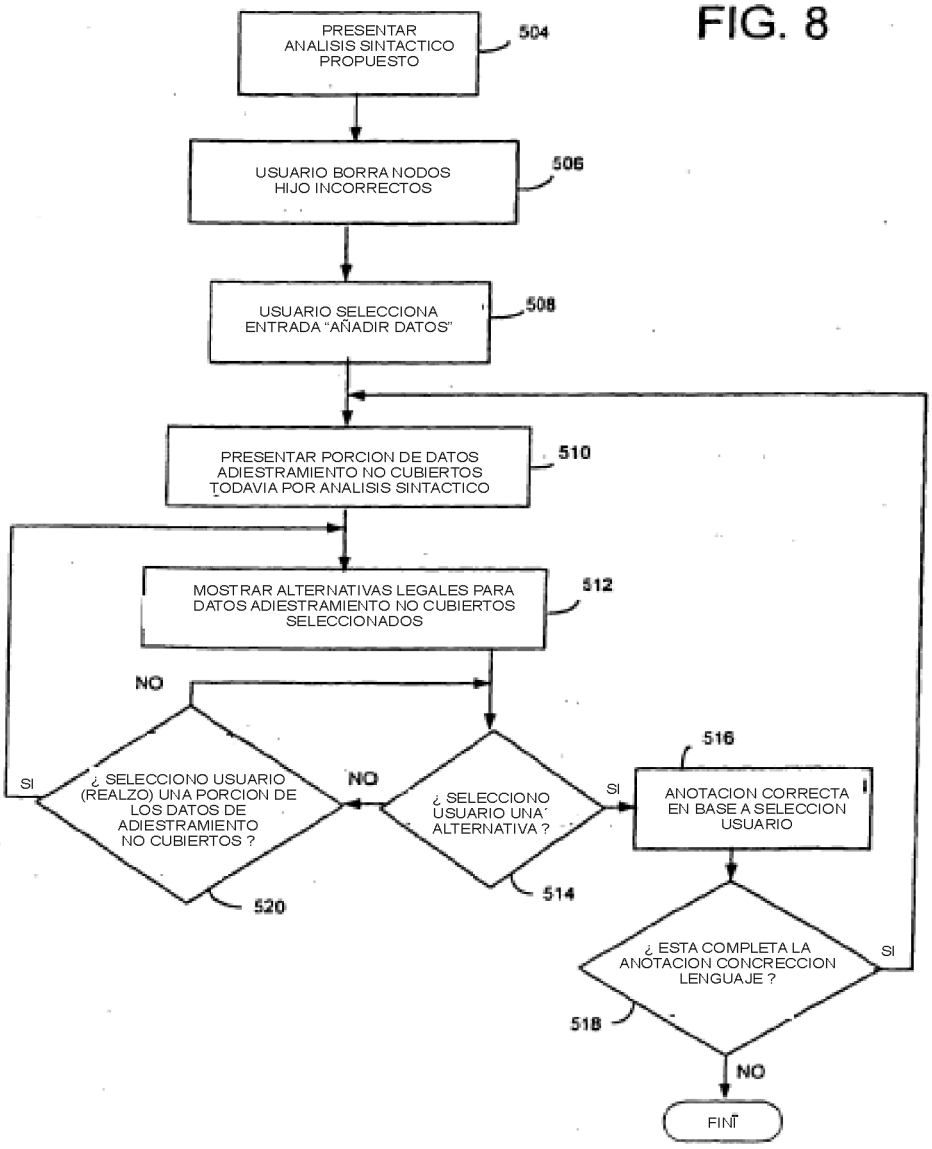

500
502
AÑADIR A ARBOL ANALISIS SINTACTICO
SEATTLE A BOSTON
CIUDAD : NOMBRE CIUDAD [5,6] - "HOUSTON"
CONECTAR: NOMBRE CIUDAD [5,6] - "HOUSTON"
NOMBRE CIUDAD LLEGADA [5,8] - "BOSTON"

FIG. 7

FIG. 8


PRESENTAR ANALISIS SINTACTICO PROPUESTO
504
USUARIO BORRA NODOS HIJO INCORRECTOS
506
USUARIO SELECCIONA ENTRADA "AÑADIR DATOS"
508
PRESENTAR PORCION DE DATOS ADIESTRAMIENTO NO CUBIERTOS TODAVIA POR ANALISIS SINTACTICO
510
MOSTRAR ALTERNATIVAS LEGALES PARA DATOS ADIESTRAMIENTO NO CUBIERTOS SELECCIONADOS
512
NO
¿ SELECCIONO USUARIO (REALZO) UNA PORCION DE LOS DATOS DE ADIESTRAMIENTO NO CUBIERTOS ?
SI
520
NO
¿ SELECCIONO USUARIO UNA ALTERNATIVA ?
514
SI
516
ANOTACION CORRECTA EN BASE A SELECCION USUARIO
¿ ESTA COMPLETA LA ANOTACION CONCRECCION LENGUAJE ?
SI
518
NO
FIN

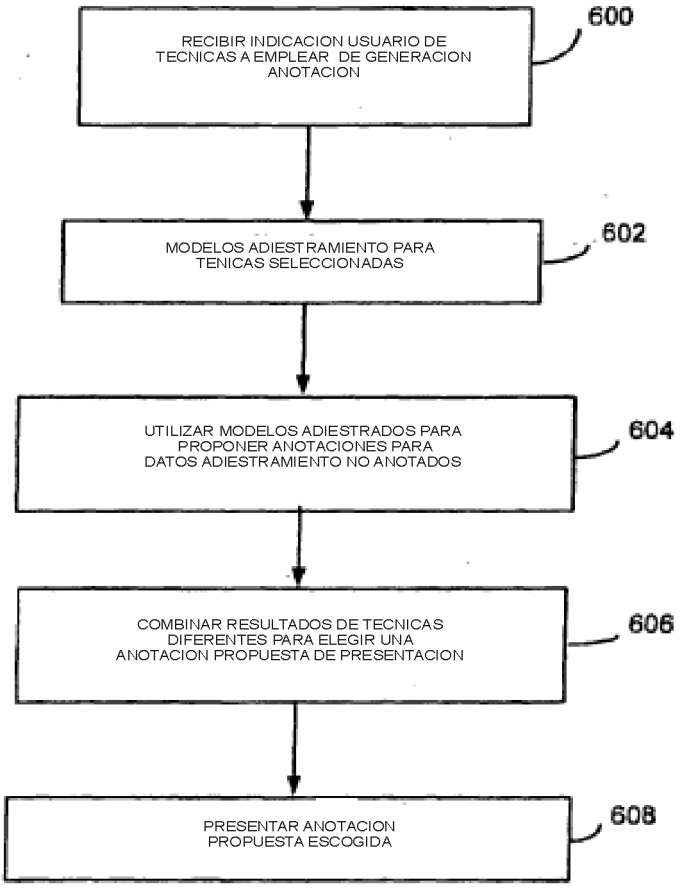

RECIBIR INDICACION USUARIO DE TECNICAS A EMPLEAR DE GENERACION ANOTACION
600
MODELOS ADIESTRAMIENTO PARA TENICAS SELECCIONADAS
602
UTILIZAR MODELOS ADIESTRADOS PARA PROPONER ANOTACIONES PARA DATOS ADIESTRAMIENTO NO ANOTADOS
604
COMBINAR RESULTADOS DE TECNICAS DIFERENTES PARA ELEGIR UNA ANOTACION PROPUESTA DE PRESENTACION
606
PRESENTAR ANOTACION PROPUESTA ESCOGIDA
608

FIG. 9