

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 946 518**

21 Número de solicitud: 202200004

51 Int. Cl.:

G06F 16/332 (2009.01)
G06F 40/30 (2010.01)
G16H 10/20 (2008.01)
G10L 15/26 (2006.01)
G10L 13/00 (2006.01)
H04L 51/02 (2012.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

19.01.2022

43 Fecha de publicación de la solicitud:

20.07.2023

71 Solicitantes:

UNIVERSIDADE DE VIGO (100.0%)
Campus Universitario de Vigo s/n
36310 Vigo (Pontevedra) ES

72 Inventor/es:

GONZÁLEZ CASTAÑO, Francisco Javier;
COSTA MONTENEGRO, Enrique;
GARCÍA MÉNDEZ, Silvia;
DE ARRIBA PÉREZ, Francisco;
JUNCAL MARTÍNEZ, Jonathan y
BURGUILLO RIAL, Juan Carlos

54 Título: **Procedimiento implementado en ordenador y sistema para analizar la capacidad de comprensión en un usuario**

57 Resumen:

Procedimiento y sistema para para analizar la capacidad de comprensión de un usuario con capacidad limitada de abstraer conceptos cuando necesita buscar información. Para ello se plantea un procedimiento y sistema que dialogue con el usuario sobre aspectos de la información que llaman su atención, todo ello mediante lenguaje natural, apoyándose en técnicas de conversión voz-texto, texto-voz y análisis de sentimiento y emoción. Un asistente conversacional conduce el diálogo sobre la información alternando preguntas dicotómicas y no-dicotómicas. Mediante técnicas de inteligencia artificial, las nuevas búsquedas se realizan combinando las palabras significativas de las respuestas del usuario con las palabras significativas contenidas en la información, en caso de que el sentimiento y emoción suscitados en el usuario sean positivos. Asimismo, mediante dichas técnicas se evalúa la capacidad de comprensión del usuario, aplicándolas al análisis de la proximidad entre sus respuestas y la información que se le presenta.

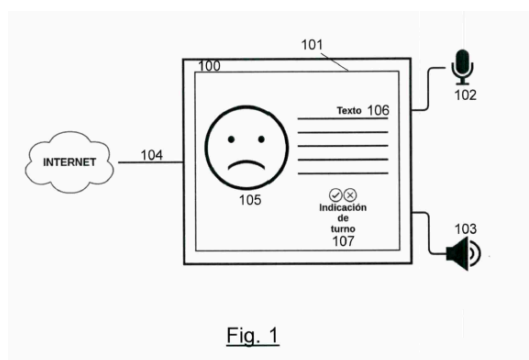


Fig. 1

DESCRIPCIÓN

Procedimiento implementado en ordenador y sistema para analizar la capacidad de comprensión en un usuario

5

Sector de la técnica

La presente invención se enmarca en el campo de los asistentes conversacionales capaces de generar, mediante inteligencia artificial, diálogos en lenguaje natural. En concreto, se refiere a técnicas para analizar la capacidad de comprensión en un usuario. A su vez, con ellas se puede estimar la capacidad cognitiva. También son útiles para consultar de forma efectiva en un buscador.

Antecedentes de la invención

15

Avances recientes en inteligencia artificial han permitido que los asistentes conversacionales, también denominados chatbots, generen respuestas de forma automática ante peticiones de los usuarios [1]. Como resultado, actualmente los asistentes conversacionales se utilizan en los negocios [2], la educación [3] y la medicina [4]. Sin embargo, imitar la inteligencia humana todavía está fuera de su alcance. Por consiguiente, los asistentes conversacionales se utilizan en tareas muy especializadas y sus interacciones son dirigidas, precisas y breves. Además, a los usuarios les resulta más fácil comunicarse con un asistente conversacional si éste emplea mensajes breves, y así permanecen interesados más tiempo [5].

25

Algunos asistentes conversacionales para entretenimiento cantan y cuentan historias. En el ámbito de la salud existen plataformas en línea para cuidado médico generalista, a través de las cuales interactúan médicos y pacientes. Ninguno de estos sistemas analiza la capacidad de comprensión, cuya estimación resulta de interés en el caso de algunos colectivos como las personas mayores. por ejemplo para la evaluación de su estado cognitivo. Aunque existen chatbots que relatan noticias. proporcionan información meteorológica o reproducen música, no se les ha dotado de empatía ni habilidades para tratar con personas con dificultades para abstraer y comprender información.

35

En lo que respecta al estado del arte previo protegido mediante patentes, aunque se ha planteado aumentar las expresiones de búsqueda mediante técnicas de aprendizaje máquina, no se ha planteado el análisis de la capacidad de comprensión de los usuarios no convencionales (como los niños autistas o los ancianos con demencia leve) mediante técnicas de inteligencia artificial. Por ejemplo, en [6] se plantea un sistema para facilitar a los ancianos la navegación en línea mediante comandos de voz, pero, en caso de problemas (como puede ser la dificultad para abstraer conceptos), el sistema se apoya en voluntarios humanos que acceden en línea.

45

La presente invención se concibe a la luz de las limitaciones del estado de la técnica. En particular, para, en base a información de cualquier tipo proporcionada a usuarios no-convencionales, esto es, con poca capacidad de abstraer conceptos, analizar de manera implícita o transparente su capacidad de comprensión, de manera más consistente y objetiva. Gracias a ello es posible disminuir los recursos de transmisión, almacenamiento y procesado de información, y por tanto dar soporte al acceso desde terminales más sencillos o baratos. puesto que, durante la duración de una sesión de usuario, se maximiza la información presentada que realmente corresponde al interés del usuario, y que por tanto resulta representativa para analizar automáticamente su grado de comprensión.

50

Referencias:

[1] Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. En Proc. Conference on Human Factors in Computing Systems, pp. 3506-3510, 2017.

[2] Thomas, N. T. (2016). An e-business chatbot using AIML and LSA. En Proc. International Conference on Advances in Computing, Communications and Informatics, pp. 2740-2742.

[3] Robins. B., Dautenhahn. K., Te Boekhorst. R., & Billard, A. (2005). Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?. Universal access in the information society, 4(2), pp.105-120.

[4] Kumar. V., & Keerthana. A. (2016). Sanative Chatbot For Health Seekers. International Journal Of Engineering And Computer Science, 5, pp. 16022-16025.

[5] Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human human online conversations and human chatbot conversations. Computers in Human Behavior 49, pp. . 245-250. [6] One-stop type intelligent comprehensive knowledge service system for the elderly based on cloud computing. Solicitud de patente en China CN103914794A.

Explicación de la invención

La presente invención se refiere a un sistema según la reivindicación independiente. Realizaciones particulares se definen en las reivindicaciones dependientes.

En general, la invención permite analizar la capacidad de comprensión de información presentada a usuarios no-convencionales, por ejemplo, usuarios con limitaciones en la capacidad de abstraer conceptos. Por ejemplo, la invención es útil para determinar el grado de interés de un usuario no-convencional en la información que consume.

Los sistemas actuales de búsqueda de información en Internet asumen que el usuario tiene capacidad de abstracción, pero algunos colectivos (personas muy mayores, personas con diversidad cognitiva) carecen de ella. Además, para estos colectivos puede ser difícil realizar búsquedas de información mediante interfaces textuales o incluso consultar la información de forma textual. El análisis automático de la capacidad de comprensión de contenidos generalistas es útil para conocer la capacidad cognitiva de un usuario y su interés real en dichos contenidos. Realizaciones de la presente invención proponen solucionar este problema planteando un sistema y un procedimiento implementado en ordenador que dialogue con el usuario no-convencional sobre información que se le presente, por ejemplo, en forma de noticias; le formule al usuario preguntas significativas relacionadas con la información (dicotómicas o no dicotómicas) generadas automáticamente; genere asimismo de forma automática respuestas válidas a dichas preguntas, y compare dichas respuestas con las del usuario para determinar su grado de comprensión; y en paralelo estime el sentimiento (positivo, negativo o neutro) y las emociones (según tipos concretos como, por ejemplo, estrés, relajación, alegría y desagrado) que expresen las respuestas del usuario (todo ello preferentemente mediante lenguaje natural, apoyándose en técnicas de conversión voz-texto y texto-voz).

Así pues, realizaciones de la presente invención permiten obtener un grado de comprensión de un usuario sobre fuentes de información en línea, de forma que, por ejemplo, se pueda evaluar

el interés real de un usuario no-convencional sobre las mismas, lo que resulta apropiado para dichos usuarios cuando experimenten dificultades para abstraer los conceptos necesarios para realizar las búsquedas. Al mismo tiempo, permitirá analizar su capacidad cognitiva.

5 Un elemento central de la invención es el uso de un asistente conversacional que induce al usuario a expresarse mediante preguntas sencillas, expresadas en voz natural. Un asistente conversacional es un programa software que interactúa con el usuario mediante voz para ofrecerle información, por ejemplo música o datos meteorológicos. Algunos ejemplos conocidos de asistentes conversacionales comerciales son Google Assistant y Alexa.

10 El asistente conversacional intercalará en la información ofrecida preguntas sencillas dirigidas al usuario basadas en la información. Estas preguntas se generan automáticamente, de acuerdo con la invención, a partir de frases clave de la información presentada que contienen palabras clave (lugares, nombres propios, etc.) u otras palabras que se estimen relacionadas con las palabras clave mediante un análisis automático de correferencias.

15 Asimismo, mediante técnicas automáticas de expansión del lenguaje se generan respuestas válidas de referencia a las preguntas partiendo de las frases clave. Una frase clave es, como se ha dicho, aquella que contiene una palabra clave o está ligada a dicha palabra clave mediante una correferencia. Por técnicas de expansión del lenguaje se entiende el subcampo de la generación de lenguaje natural que permite crear frases sintáctica y semánticamente coherentes a partir de un conjunto de palabras.

20 Además, se aplican a las respuestas del usuario análisis de sentimiento y emoción basados en técnicas de inteligencia artificial, que se detallan más adelante. En caso de que el sentimiento sea positivo o las emociones pertenezcan a categorías específicas (por ejemplo, alegría), las palabras clave contenidas en la respuesta del usuario (por ejemplo, sustantivos, adjetivos y verbos) y las palabras clave de la información (por ejemplo, lugares y nombres propios) se emplean para buscar en Internet la siguiente información que se presentará al usuario.

30 Finalmente, se comparan mediante una métrica en un mismo espacio numérico, como se detalla más adelante, las respuestas del usuario con las respuestas de referencia generadas mediante la invención, y, dependiendo de la distancia a la que se encuentren o, indirectamente, la aplicación de clasificadores, se estima la capacidad de comprensión del usuario.

35 Así pues, una mayor distancia es una medida de la disparidad (poca proximidad) semántica entre el contenido presentado y lo expresado por el usuario, que permite evaluar su capacidad de comprensión. Se mide entre los vectores fruto de la conversión de los textos a comparar, especialmente las palabras con carga semántica presentes en la información y las respuestas, a un mismo espacio numérico multidimensional.

40 La distancia se puede pesar de forma diferenciada por componentes tales como n-gramas, frecuencia de palabras, número de palabras de determinado tipo (sustantivos, adjetivos, verbos, etc.) o cualquiera de los métodos habituales en la disciplina del procesamiento de lenguaje natural.

45 Es decir, el propio diálogo generado se utiliza para evaluar la capacidad de comprensión de la información presentada y deducir los intereses de un usuario no-convencional de cara a realizar búsquedas.

50 Gracias a la invención es posible disminuir los recursos de transmisión, almacenamiento y procesamiento de información, y por tanto dar soporte al acceso desde terminales más sencillos o baratos, puesto que, durante la duración de una sesión de usuario, se maximiza la información

presentada que realmente corresponde al interés del usuario y que por tanto resulta representativa para analizar automáticamente su grado de comprensión.

Breve descripción de las figuras

- 5
Figura 1 es un ejemplo muy simplificado de un sistema (100) de acuerdo con la invención.
- Figura 2 es un ejemplo más detallado de un diagrama con bloques funcionales de un sistema (100) de acuerdo con la invención.
- 10
Figura 3 es un ejemplo de diagrama de flujo de acuerdo con la invención.

GLOSARIO DE REFERENCIAS NUMÉRICAS

- 15 100 Sistema
- 101 Pantalla
- 102 Micrófono
- 20 103 Altavoz
- 105 Avatar
- 25 107 Icono de indicación de turno de palabra
- 201 Información presentada
- 202 Conversar de texto a voz
- 30 203 Conversar de voz a texto
- 204 Analizador de sentimiento y emoción
- 35 205 Asistente conversacional
- 206 Generador de lenguaje natural
- 207 Segundo extractor de palabras clave
- 40 208 Analizador de contenido
- 209 Buscador
- 45 210 Extractor de contenido
- 211 Primer extractor de palabras clave
- 212 Base de datos de prototipos

213 Motor de diálogo

301t Primer mensaje de texto

5 301v Primer mensaje de voz 311 Contenido

311 b Búsqueda o consulta

311p Palabras clave del mensaje de texto

10

302t Segundo mensaje de texto

302v Segundo mensaje de voz

15 320 Usuario

320c Grado de comprensión del contenido

321 Primera respuesta del usuario

20

321p Palabras clave de la primera respuesta del usuario

321r Respuesta de referencia

25 **Realización preferente de la invención**

En una realización, la Figura 1 muestra algunos elementos de un sistema (100) de acuerdo con la invención. En una pantalla (101) se muestra una representación simbólica de un personaje que refleja emociones, como un avatar (105). En el ejemplo de la realización el personaje refleja tristeza. En la pantalla (101) se muestra un texto que el sistema (100) reproduce como voz sintética por el altavoz (103). En la pantalla (101) se muestra un icono de indicación de turno de palabra (107) con dos estados. Uno de los estados le indica al usuario cuando puede hablar de manera que un micrófono (102) capte la voz del usuario. El otro indica que no debe hacerlo.

30

En una realización, el sistema dirige el diálogo mostrando en pantalla un icono verde cuando se espera una respuesta del usuario y rojo cuando el propio sistema está hablando, o indica los turnos con dos iconos distintos cualesquiera.

35

En una realización, cuando el sistema está hablando muestra en pantalla el texto transcrito (106) de su locución.

40

En una realización, para determinado sentimiento o determinadas emociones del usuario o la locución del dispositivo, el avatar (105) muestra una expresión acorde mediante las convenciones gráficas usuales de codificación de sentimiento, tales como emoticonos.

45

En una realización, la Figura 2 muestra los bloques funcionales que intervienen en algunas funciones inteligentes de un sistema (100) de acuerdo con la invención.

El asistente conversacional (205) proporciona al usuario preguntas dicotómicas y no dicotómicas basadas en la información presentada (201), que se transforman en voz mediante un conversor

50

texto a voz (202). Un extractor de contenido (210) obtiene la información presentada con técnicas como por ejemplo "scrapping" aplicadas a los resultados de un buscador (209) de Internet. Un generador de lenguaje natural (206) genera automáticamente las preguntas. El generador de lenguaje natural (206) identifica los nombres propios de personas, lugares y organizaciones en la información presentada, contrastando las palabras de la información en bases de datos. En una realización, la generación de las preguntas tiene lugar como sigue:

En el caso de las preguntas dicotómicas, orientadas al establecimiento inicial de la conversación y la obtención de preferencias de los usuarios sobre el contenido, se generan a partir de plantillas tomadas de una base de datos (212) con prototipos de cuestiones en las que se combinan los nombres propios con un conjunto de palabras que doten de significado y coherencia al conjunto de la frase. Por ejemplo, si la información habla de 'Berlín', las preguntas generadas podrían ser "¿Has estado alguna vez en Berlín?" o "¿Te interesan noticias relacionadas con la ciudad de Berlín?". Por otro lado, si la información habla de una persona llamada "Inés" las preguntas generadas podrían ser "¿Conoces a Inés?". "¿Te ha gustado escuchar información de esta persona?"

En el caso de las preguntas no dicotómicas, se detectan en la información presentada los sintagmas nominales de las oraciones, que contienen los nombres propios, y los verbos que los acompañan. Adicionalmente, se aplican técnicas de correferencia que relacionan ciertos pronombres o sintagmas nominales con el sustantivo al que hacen referencia. Por ejemplo, en "París es una ciudad preciosa. Esta ciudad se sitúa en el centro de Francia", "esta ciudad" hace referencia a París. Los sintagmas nominales fruto del análisis de correferencias y los verbos que los acompañan, en consecuencia, también se consideran detectados. Por ejemplo, seleccionado un sintagma nominal detectado y su verbo, si ambos están acompañados por una fecha, como "12 de febrero", se genera la pregunta: "¿Qué dice la noticia que pasa el 12 de febrero?", cuya respuesta correcta sería el propio sintagma nominal. Si un sintagma nominal, por ejemplo "4 kilos de droga", contiene números y texto, se separa el texto de los números y se construye una pregunta del tipo "¿Cuántos kilos de droga se han incautado?" cuya respuesta sería el valor "4". En otro ejemplo, si se detecta un lugar o persona como sintagma nominal de una oración, se construye una pregunta utilizando las partículas interrogativas qué, quién/quiénes, dónde, seguidas del verbo que acompaña al sintagma nominal. Por ejemplo, "¿Quién es el campeón de Fórmula Uno?", cuya respuesta sería el sintagma nominal "Verstappen". Adicionalmente, se pueden construir tipos adicionales de preguntas de respuesta no dicotómica como: "¿Qué lugares menciona la noticia?" cuya respuesta es una lista de los lugares identificados y "¿De cuántas personas habla la noticia?" o "¿De cuántos lugares habla la noticia?", donde las respuestas correctas vienen dada por los números de personas y lugares detectados en el texto.

En el caso de las preguntas dicotómicas, según una realización, no existe una respuesta correcta para el sistema (100). tan solo permiten que el usuario exprese un sentimiento o una emoción, o recabar su opinión. En el caso de las preguntas no dicotómicas, el generador de lenguaje natural (206) también puede generar automáticamente respuestas correctas para todas las preguntas. Consisten en el sintagma nominal, eliminando las palabras irrelevantes conocidas como palabras sin valor semántico o de parada.

Las respuestas del usuario se traducen a texto mediante un conversor voz a texto (203) y el texto resultante se somete a un analizador de sentimiento y emoción (204). En una realización, el analizador de sentimiento y emoción (204) puede utilizar técnicas de aprendizaje máquina supervisado. Por ejemplo, mediante clasificadores entrenados utilizando máquinas de vectores soporte, árboles de decisión, K vecinos más próximos o cualquier otro, donde se parte de un conjunto de textos de referencia (interacciones pregrabadas con usuarios, textos tomados de redes sociales o cualesquiera otros), que un anotador humano categoriza en base al sentimiento (negativo, neutro o positivo) o emoción de un conjunto dado (por ejemplo alegría, interés,

optimismo, desagrado, estrés y desconcierto) que a su juicio dichos textos expresan. Tras ello, se entrena uno de los clasificadores antedichos o un conjunto de los mismos, en serie o paralelo. para obtener un resultado similar, al aplicarlo a una conversión del texto de entrada al mismo espacio numérico multidimensional en el que se realizó el entrenamiento del modelo.

5 Alternativamente, el analizador de sentimiento y emoción (204) puede utilizar técnicas no supervisadas. basadas en bolsas de palabras o propagación de valores de relevancia de palabras a través de árboles sintácticos, o cualquiera otra. Si el número de respuestas positivas o de detecciones de ciertas emociones (por ejemplo, alegría e interés) supera un umbral configurable, el analizador de contenido (208) puede suponer que una información ha sido de

10 interés para el usuario.

El motor de diálogo (213) establece el orden de presentación de informaciones, preguntas dicotómicas y preguntas no dicotómicas, de manera que cualquiera de estos elementos puede

15 repetirse el número de veces que se estime conveniente. En una realización, el motor de diálogo (213) emplea la secuencia de tipos de preguntas de un test de evaluación cognitiva. por ejemplo una fase breve de preguntas dicotómicas seguidas de varias preguntas no dicotómicas con orientaciones específicas a palabras clave. El motor de diálogo (213) puede alterar esta estructura en función del sentimiento y emoción detectados por el analizador de sentimiento y

20 emoción (204), por ejemplo, si el número de respuestas del usuario con sentimiento negativo excede un umbral.

A fin de introducir un sentimiento de empatía y de reconocimiento del interlocutor, el generador de lenguaje natural (206) puede ajustar las preguntas que genera incorporando en la generación

25 de nuevas preguntas palabras clave (como sustantivos, verbos y adjetivos) de las respuestas del usuario y el sentimiento y emoción detectados en el usuario (204). Así, por ejemplo, en el caso de las preguntas no dicotómicas, si se detecta en una respuesta del usuario el lugar "Madrid" y un sentimiento negativo. la siguiente pregunta no dicotómica podría ser "¿Por qué no te gusta Madrid?". Para dotar al asistente conversacional (205) de la flexibilidad que no proporcionan las

30 plantillas, en una realización, el generador de lenguaje natural (206) puede apoyarse para la generación de las preguntas en un expansor de lenguaje natural, que, partiendo de un conjunto de palabras clave tomadas de la información presentada y las respuestas del usuario. el contexto de la aplicación (como hora, tiempo meteorológico. datos personales del usuario, etc.) y el sentimiento y emoción detectados en el usuario por el analizador de sentimiento y emoción (204)

35 o cualquier otro dato, genere texto coherente. Por ejemplo, el expansor puede utilizar, aparte de las plantillas antedichas, conocimiento lingüístico (léxicos, gramáticas como árboles sintácticos) o conocimiento estadístico. como las metodologías BERT, BART o GPT.

A fin de evaluar la capacidad de comprensión del usuario, el analizador de contenido (208)

40 calcula la proximidad semántica entre las palabras clave de las expresiones del usuario (proporcionadas por el extractor de palabras clave (207)) y la información presentada al usuario (201).

El mismo analizador de contenido (208) permite realizar una abstracción automática de los

45 intereses del usuario para ofrecerle nueva información, decidiendo por el usuario las búsquedas de nueva información mediante un buscador (209) de Internet. Esta funcionalidad puede ser útil en diversas circunstancias. y facilita el acceso a las nuevas tecnologías de forma más interactiva para el usuario y más eficiente, ya que se emplea menos tiempo de conexión y se desperdicia menos ancho de banda para acceder a información deseada. El analizador de contenido (208)

50 prioriza dos tipos de palabras clave. Las palabras clave del contenido extraído de Internet, proporcionadas por el primer extractor de palabras clave (211), y las palabras clave extraídas de las respuestas del usuario por el segundo extractor de palabras clave (207) cuando el analizador de sentimiento y emoción (204) detecta que el número de respuestas positivas y respuestas con

ciertas emociones (por ejemplo alegría) superan cierto umbral y existe proximidad semántica entre las palabras clave de las respuestas del usuario y las de la información presentada. En una realización, el primer y segundo extractor de palabras clave pueden ser el mismo.

5 A fin de definir la proximidad semántica entre dos textos, el analizador de contenido (208) puede proceder como sigue, definiendo una métrica de la proximidad semántica. Aunque el desarrollo matemático que sigue debe considerarse un ejemplo y podría sustituirse por cualquier otra medida de proximidad en un espacio numérico común al que se convierten los textos a comparar:

10 En primer lugar, se deben identificar las palabras clave en ambos textos, por ejemplo, sustantivos, adjetivos, verbos y adverbios. Utilizando bases de datos lingüísticas se pueden conocer los holónimos, hiperónimos, hipérnimos, hipónimos, merónimos y sinónimos de dichas palabras clave.

15 Acto seguido se puede asignar una puntuación de proximidad a una expresión del usuario como sigue:

$$sim = 0.8 \frac{\sum_{i=1}^{N_{sust}} sust_i + \sum_{i=1}^{N_{verb}} verb_i}{N_{sust} + N_{verb}} + 0.2 \frac{\sum_{i=1}^{N_{adj}} adj_i + \sum_{i=1}^{N_{adv}} adv_i}{N_{adj} + N_{adv}}$$

Dónde:

- 20 • N_x representa el número de palabras de la categoría léxica x (sustantivos *sust*, verbos *verb*, adjetivos *adj* o adverbios *adv*) en una respuesta ideal que contiene las palabras clave de la información presentada.
- 25 • Dada la palabra clave i -ésima de la categoría x en la respuesta ideal, se calcula su similaridad con todas las palabras de la misma categoría léxica x en la respuesta del usuario y se toma el valor máximo como x^*i .

30 La expresión anterior debe considerarse un ejemplo que considera que los sustantivos y los verbos son más relevantes que adjetivos y adverbios, debido a los factores multiplicadores respectivos 0.8 y 0.2. Esto es un ejemplo sin pérdida de generalidad, puesto que se puede asignar cualquier factor multiplicador que se considere conveniente a cualquier categoría. En general, no obstante, verbos y adjetivos suelen poseer mayor información semántica.

35 Para obtener x^*i se puede estimar la similaridad $s(pal1, pal2)$ entre las palabras $pal1$ y $pal2$ como sigue:

$$s(pal1, pal2) = (1 - \gamma)\alpha\beta^{d(pal1, pal2)} + \gamma$$

40 Dónde:

$\alpha=0.9$ si las palabras son sinónimos y 0.85 en otro caso.

45 $d(pal1, pal2) = 0$ si las palabras son holónimos, hipérnimos, hipónimos, merónimos o sinónimos, relacionados con un mismo dominio de significado de una base de datos lingüística. En otro caso, $d(pal1, pal2)$ se toma como el camino más corto entre $pal1$ y $pal2$ dentro de la estructura jerárquica de una base de datos lingüística, como por ejemplo Wordnet.

β es un factor de profundidad que decrementa la similaridad de modo exponencial, como una potencia al número de pasos que separan las dos palabras como transformaciones dentro de una base de datos lingüística. Por ejemplo $\beta=0.7$.

- 5 γ es una constante que se puede ajustar mediante prueba y error para obtener el mejor resultado posible.

Los valores numéricos planteados son ejemplos válidos para información de tipo noticias pero deben considerarse ejemplos y pueden sustituirse por otros cualesquiera.

- 10 En una realización, el analizador de contenido (208) puede seleccionar las palabras del usuario que pertenezcan a la misma categoría semántica que las palabras clave de la información presentada (excluyendo las palabras con nula o baja carga semántica) mediante la herramienta Multilingual Central Repository.

- 15 Un extractor de contenido (210) extrae las noticias de los sitios de Internet que obtiene el buscador (209). En una realización, el extractor de noticias es un "scraper" que extrae texto de páginas web de periódicos, portales, blogs o noticieros.

- 20 En una realización, para generar preguntas no dicotómicas con carga de evaluación se emplea la base de datos (212). con prototipos de preguntas tomados de tests de evaluación cognitiva tipo CAMDEX. El motor de diálogo puede decidir el instante adecuado para utilizar un prototipo u otro en función del sentimiento o emoción del usuario o su grado de comprensión, estimados respectivamente por el analizador de sentimiento y emoción (204) y el analizador de contenido (208). en el segundo caso a partir de la proximidad entre la información presentada (201) y las respuestas del usuario o palabras clave extraídas (207) de las mismas, dependiendo de sus sentimientos y emociones proporcionados por el analizador de sentimiento y emoción (204) como se ha explicado previamente, bien directamente (contrastando la proximidad frente a una escala numérica), o indirectamente, procesando la proximidad y el sentimiento y emociones detectados a través de un clasificador supervisado de grados de comprensión. entrenado mediante un modelo como máquinas de vectores soporte, árboles de decisión, K vecinos más próximos, o cualquier otro.

- 35 La Figura 3 ilustra un ejemplo de procedimiento para analizar la capacidad de comprensión en un usuario mediante un diagrama de flujo con distintos pasos llevados a cabo por componentes funcionales mencionados en la realización anterior.

- 40 El procedimiento permite analizar la capacidad de comprensión en un usuario (320). Para ello, se extrae un contenido (311) de una fuente de información de Internet. La fuente de información es típicamente un documento. por ejemplo, un archivo PDF. o una página web de un periódico. Esta tarea la realiza un extractor de contenido (210).

- 45 A continuación, se extraen palabras clave (311p) del contenido (311) extraído. Esta tarea la realiza un primer extractor de palabras clave (211).

- Seguidamente, se genera un primer mensaje de texto (301t) usando una o más palabras clave (311p). Esta tarea la realiza un generador de lenguaje natural (206). El primer mensaje de texto (301t) incluye una pregunta.

- 50 Una vez, generado el primer mensaje de texto (301t) se convierte en un primer mensaje de voz (301v). De esta tarea se encarga un conversor de texto a voz (202). Es de esperar que el usuario

(320), al escuchar el primer mensaje de voz, emita una primera respuesta (321), que se recoge y transcribe. De esta tarea se encarga un conversor de voz a texto (203).

Entonces, entra en juego un segundo extractor de palabras clave (207), que puede ser el mismo extractor de palabras clave (211) que se usó antes. Así, se extrae un conjunto de palabras clave (321p) de la primera respuesta (321).

Prosigue, con una generación de una respuesta de referencia (321r). De esta tarea también se encarga el generador de lenguaje natural (206), aplicando técnicas de expansión del lenguaje al primer mensaje de texto (3010).

Sucede una comparación entre la respuesta de referencia (321r) y el conjunto de palabras clave (321p) de la primera respuesta (321). Para ello, se aplica una métrica de proximidad semántica entre palabras, como la descrita en el ejemplo previo de la Figura 2. Un analizador de contenido (208) realiza esta función.

Finalmente, con el mencionado analizador de contenido (208), se asigna un grado de comprensión (320c) del contenido (311) en función de la métrica de proximidad semántica.

Opcionalmente, se contempla una realización en la que se puede analizar con un analizador de sentimiento y emoción (204) si la respuesta (321) del usuario presenta un sentimiento positivo, negativo o neutro. Adicionalmente, o alternativamente, se puede analizar si la respuesta (321) presenta un tipo de emoción concreta como estrés, relajación, alegría y desagrado. Gracias a la información recogida del usuario, el analizador de contenido (208) puede conectarse a un buscador (209) y realizar una consulta o búsqueda (311b) en Internet de una manera automática, más enriquecida y personalizada al usuario y momento concretos. Para ello, la consulta o búsqueda (311b) incorpora una o más palabras clave (311p) del primer mensaje de texto (301t), si el sentimiento y/o emoción de? usuario (320) son indicativos de interés respecto del contenido (311) del que proviene el primer mensaje de texto (301t) extraído en el inicio. Esta realización puede implementarse como un asistente de búsqueda interactivo. Los resultados que ofrece el buscador (209) son más personalizados y ahorran tiempo al usuario. Especialmente, si el usuario tiene dificultades con las nuevas tecnologías. También, permite gestionar más convenientemente el uso de recursos técnicos (ancho de banda, consumo de datos, batería, etc.), puesto que se sirve el contenido de interés en menor plazo.

La descripción y los dibujos ilustran los principios de la invención. Aunque la presente invención se ha descrito con referencia a realizaciones específicas, los expertos en la materia deben entender que tanto dichas modificaciones como sus detalles pueden sufrir cambios sin alejarse del alcance de la presente invención como definen las reivindicaciones.

Además, todos los ejemplos mencionados en el presente documento están principalmente y expresamente destinados a fines pedagógicos para ayudar al lector a comprender los principios de la invención y los conceptos aportados por el o los inventores para fomentar la técnica, y deben interpretarse sin limitación a los ejemplos y condiciones específicamente mencionados. Además, todas las afirmaciones del presente documento que mencionan principios, aspectos y realizaciones de la invención, así como sus ejemplos específicos, pretenden abarcar sus equivalentes.

El procedimiento y el sistema presentes se pueden implementar en un dispositivo electrónico de cualquier tipo, con capacidad de procesamiento, como un teléfono móvil, un portátil, un ordenador de sobremesa o un dispositivo fabricado ex profeso, con acceso a una pantalla gráfica (101), un micrófono (102), un altavoz (103) y a fuentes de información en línea, como Internet

(104), que ejecuta un asistente conversacional inteligente capaz de sostener diálogos con un usuario para ayudarlo a abstraer conceptos de cara a realizar búsquedas de información y, en base a las interacciones con el usuario, evaluar su grado de comprensión. Para ello, el asistente inteligente se apoya en una serie de elementos funcionales que le proporcionan funciones de inteligencia artificial y procesamiento de lenguaje natural.

En algunas realizaciones, algunos de los periféricos hardware antedichos no estarán incluidos en el sistema, sino que serán externos, o el sistema se conectará con dispositivos que dispongan de dichos periféricos para utilizarlos cuando los precise.

10

REIVINDICACIONES

1. Procedimiento implementado en ordenador para analizar la capacidad de comprensión en un usuario (320), comprendiendo dicho procedimiento los pasos de:

- 5 - extraer un contenido (311) de una fuente de información de Internet, donde la fuente de información de Internet es un documento o una página web, mediante un extractor de contenido (210);
- 10 - extraer una pluralidad de palabras clave (311p) del contenido (311), mediante un primer extractor de palabras clave (211);
- generar un primer mensaje de texto (301t) con una o más palabras clave (311p) extraídas, mediante un generador de lenguaje natural (206), donde el primer mensaje de texto (301t)
15 comprende una pregunta;
- convertir el primer mensaje de texto (301t) en un primer mensaje de voz (301v), mediante un conversor de texto a voz (202);
- 20 - recoger y transcribir una primera respuesta (321) del usuario (320) tras el primer mensaje de voz, mediante un conversor de voz a texto (203);
- extraer un conjunto de palabras clave (321p) de la primera respuesta (321), mediante un
25 segundo extractor de palabras clave (207);
- generar una respuesta de referencia (321r), mediante el generador de lenguaje natural (206), aplicando técnicas de expansión del lenguaje al primer mensaje de texto (301 t);
- comparar la respuesta de referencia (321r) y el conjunto de palabras clave (321p) de la primera
30 respuesta (321), aplicando una métrica de proximidad semántica entre palabras, mediante un analizador de contenido (208);
- asignar un grado de comprensión (320c) del contenido (311) en función de la métrica de
35 proximidad semántica, mediante el analizador de contenido (208).

2. Procedimiento según la reivindicación 1, que comprende además:

- analizar si una primera respuesta (321) del usuario presenta un sentimiento positivo, negativo
40 o neutro, mediante un analizador de sentimiento y emoción (204);
- analizar si la primera respuesta (321) presenta un tipo de emoción concreta que comprende estrés, relajación, alegría, tristeza, agrado y desagrado, mediante un analizador de sentimiento y emoción (204);
- 45 - en función del sentimiento y del tipo de emoción presentados en la primera respuesta (321), realizar una búsqueda (311b) en Internet, mediante un buscador (209), donde la búsqueda comprende palabras clave del primer mensaje de texto (311 p) y/o la respuesta del usuario (321 p).
- 50 3. Procedimiento según la reivindicación 1 o 2, donde la pregunta del primer mensaje de texto (3011) es no-dicotómica, donde el procedimiento comprende además los pasos de:

- 5 - generar un segundo mensaje de texto (302t), donde el segundo mensaje de texto (302) comprende una pregunta dicotómica con una o más palabras clave del contenido (311), donde la pregunta dicotómica se genera a partir de una plantilla en una base de datos (212) o a partir de un expansor de lenguaje natural alimentado con palabras clave del contenido (311) y/o de la respuesta (321) del usuario (320);
- 10 - convertir el segundo mensaje de texto (302t) en un segundo mensaje de voz (302v), mediante un conversor de texto a voz (202);
- recoger y transcribir una segunda respuesta (322) del usuario (320) tras el segundo mensaje de voz, mediante un conversor de voz a texto (203);
- 15 - generar una respuesta de referencia (321 r) a la pregunta dicotómica;
- comparar la respuesta de referencia (321 r) y la transcripción de la segunda respuesta (322);
- modificar el grado de comprensión (320c) del contenido (311) asignado en función de la comparación, mediante el analizador de contenido (208).
- 20 4. Procedimiento según una cualquiera de las reivindicaciones anteriores, que comprende además estimar un grado cognitivo para el usuario a partir de una pluralidad de asignaciones de grados de comprensión, cada una asociada a un contenido (311) diferente y/o a preguntas dicotómicas y no-dicotómicas diferentes.
- 25 5. Sistema para analizar la capacidad de comprensión en un usuario (320), que comprende:
 - un extractor de contenido (209) configurado para extraer un contenido (311) de una fuente de información de Internet, donde la fuente de información de Internet es un documento o una página web;
 - un primer extractor de palabras clave (211) configurado para extraer una pluralidad de palabras clave del contenido (311);
 - 35 - un generador de lenguaje natural (206) configurado para generar un primer mensaje de texto (3011) con una o más palabras clave, donde el primer mensaje de texto (3011) comprende una pregunta;
 - un conversor de texto a voz (202) configurado para convertir el primer mensaje de texto (3011) en un primer mensaje de voz (301 v);
 - 40 - un conversor de voz a texto (203) configurado para recoger y transcribir una primera respuesta (321) del usuario (320) tras el primer mensaje de voz;
 - 45 - un segundo extractor de palabras clave (207) configurado para extraer un conjunto de palabras clave (321 p) de la primera respuesta (321); donde el generador de lenguaje natural (206) está además configurado para generar una respuesta de referencia (321 r), aplicando técnicas de expansión del lenguaje, al primer mensaje de texto (3011);
 - 50 - un analizador de contenido (208) configurado para comparar la respuesta de referencia (321 r) y el conjunto de palabras clave (321 p) de la primera respuesta (321), a partir una métrica de

proximidad semántica entre palabras, donde el analizador de contenido (208) está además configurado para asignar un grado de comprensión (320c) del contenido (311) en función de la métrica de proximidad semántica.

5 6. Sistema según la reivindicación 5, que comprende además un analizador de sentimiento y emoción (204) configurado para:

- analizar si una primera respuesta (321) del usuario presenta un sentimiento positivo, negativo o neutro;

10 - analizar si la primera respuesta (321) presenta un tipo de emoción concreta que comprende estrés, relajación, alegría, tristeza, agrado y desagrado;

15 - conectarse a un buscador (209) y realizar una búsqueda (311b) en Internet, en función del sentimiento y del tipo de emoción presentados en la primera respuesta (321), donde la búsqueda comprende palabras clave del primer mensaje de texto (301t).

20 7. Sistema según la reivindicación 5 o 6, donde la pregunta del primer mensaje de texto (301t) es no-dicotómica, donde el generador de lenguaje natural (206) está configurado además para generar un segundo mensaje de texto (302t), donde el segundo mensaje de texto (302t) comprende una pregunta dicotómica con una o más palabras clave del contenido (311), donde la pregunta dicotómica se genera accediendo a una plantilla en una base de datos (212) o a partir de un expansor de lenguaje natural alimentado con palabras clave del contenido (311) y/o de la respuesta (321) del usuario (320);

25 donde el conversor de texto a voz (202) está configurado para convertir el segundo mensaje de texto (302t) en un segundo mensaje de voz (302v);

30 donde el conversor de voz a texto (203) está configurado para recoger y transcribir una segunda respuesta (322) del usuario (320) tras el segundo mensaje de voz;

donde el generador de lenguaje natural está configurado para generar una respuesta de referencia (321 r) a la pregunta dicotómica;

35 donde el analizador de contenido (208) está configurado para:

comparar la respuesta de referencia (321 r) y la transcripción de la segunda respuesta (322) y para modificar el grado de comprensión (320c) del contenido (311) asignado en función de la comparación.

40 8. Sistema según una cualquiera de las reivindicaciones anteriores 5 a 7, donde el analizador de contenido (208) está configurado además para estimar un grado cognitivo para el usuario a partir de una pluralidad de asignaciones de grados de comprensión, cada una asociada a un contenido (311) diferente y/o a preguntas dicotómicas y no-dicotómicas diferentes.

45 9. Medio legible por ordenador que comprende un conjunto de instrucciones codificadas para ser ejecutadas y llevar a cabo los pasos del método según una cualquiera de las reivindicaciones anteriores 1 a 4.

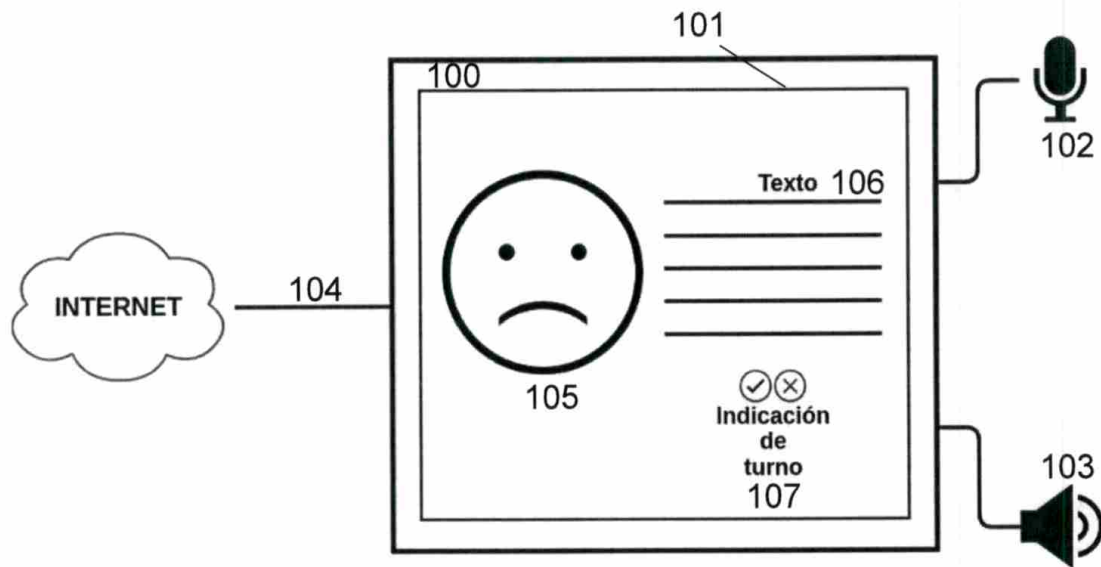


Fig. 1

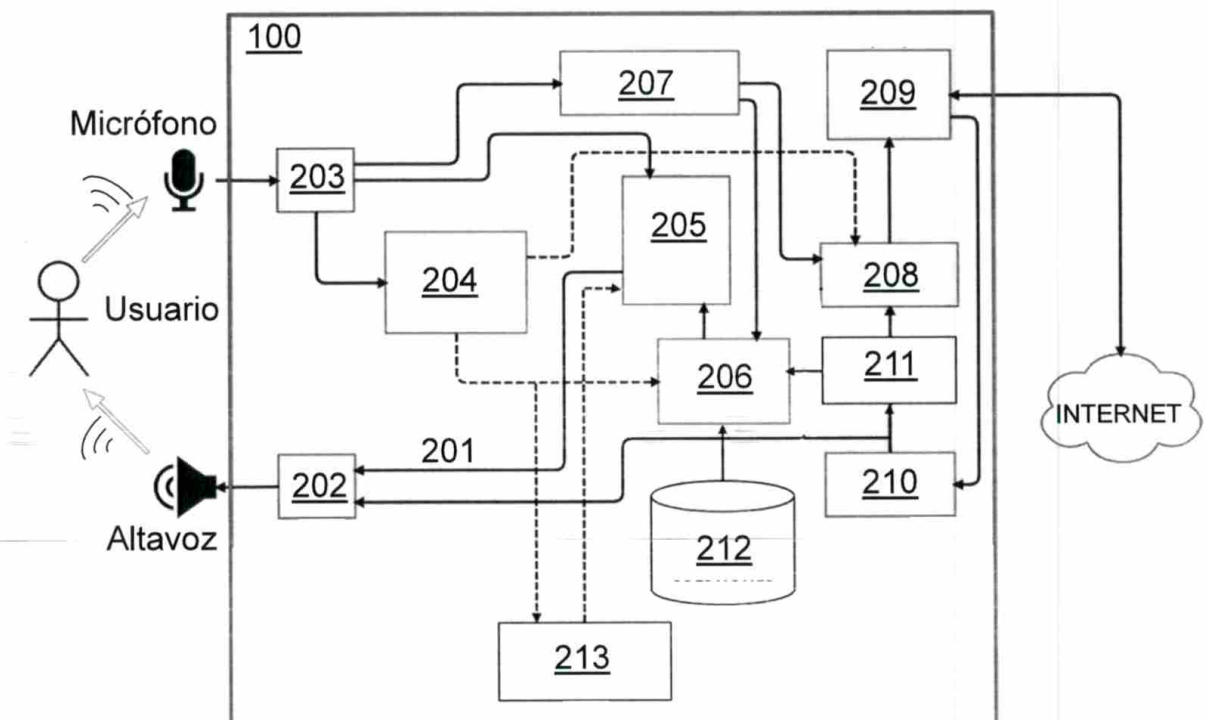


Fig.2

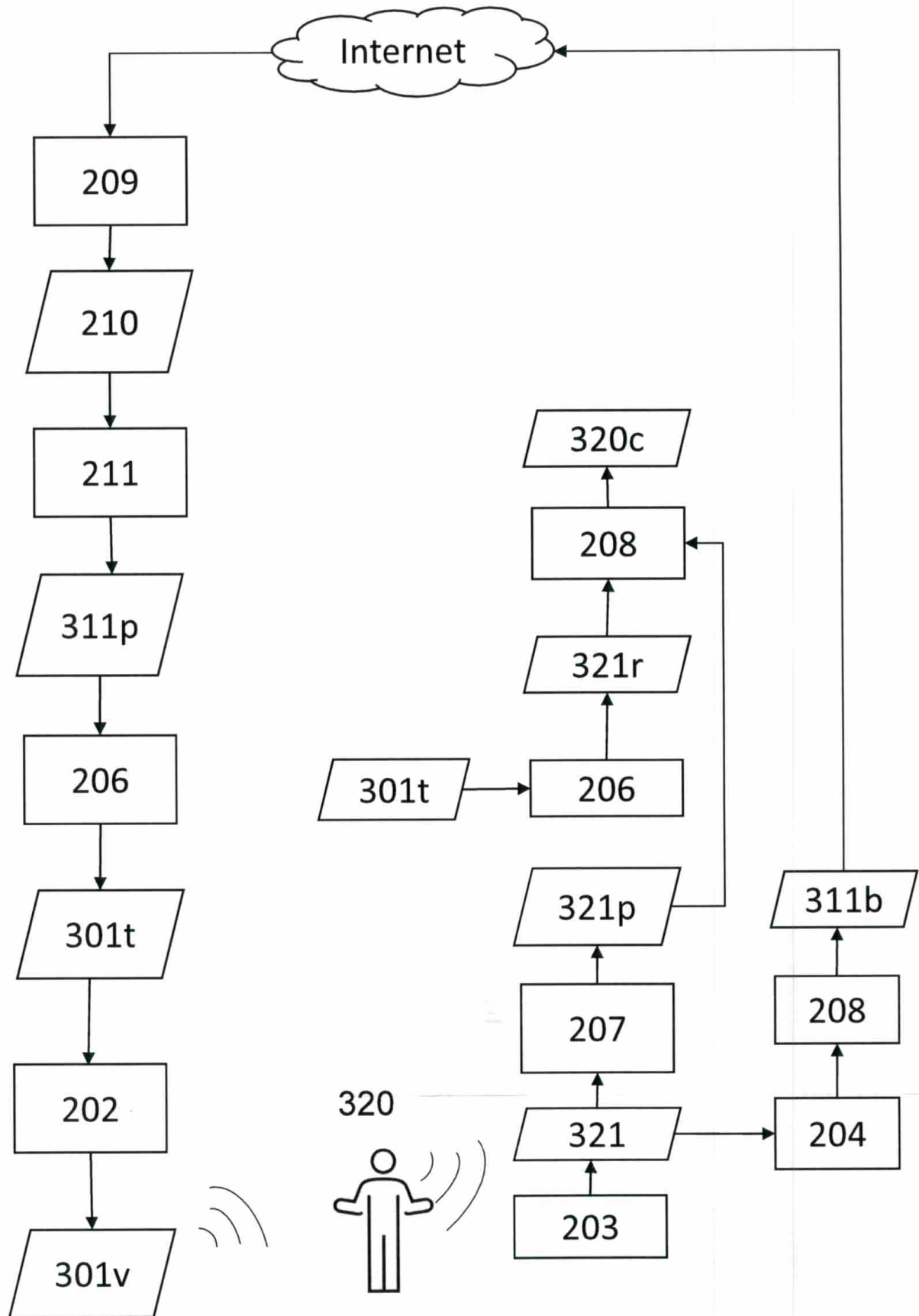


Fig. 3



- ②① N.º solicitud: 202200004
②② Fecha de presentación de la solicitud: 19.01.2022
③② Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤① Int. cl.: Ver Hoja Adicional

DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
X	GARCIA-MENDEZ, SILVIA et al.: "Entertainment Chatbot for the Digital Inclusion of Elderly People Without Abstraction Capabilities". IEEE Access, 20210517 IEEE, USA, 17/05/2021, Vol. 9, Páginas 75878 - 75891 [en línea] [recuperado el 02/09/2022]. Recuperado de Internet <URL: https://ieeexplore.ieee.org/abstract/document/9432816 >, <DOI: doi:10.1109/ACCESS.2021.3080837>. Apartados III, III.A, III.B, III.C, III.D.3, III.D.5, V; figuras 1-2, 5; tablas 1-2.	1-9
L	KR 101980348B B1 (KIM JEONG HYU) 20/05/2019, párrafos 1, 6, 11, 19-21, 28, 32, 36-65, 76-79, 88-96, 110-117, 133-135; figuras 1-5.	1-9
Categoría de los documentos citados X: de particular relevancia Y: de particular relevancia combinado con otro/s de la misma categoría A: refleja el estado de la técnica O: referido a divulgación no escrita P: publicado entre la fecha de prioridad y la de presentación de la solicitud E: documento anterior, pero publicado después de la fecha de presentación de la solicitud L: documento que puede plantear dudas sobre una reivindicación de prioridad o que se cita para determinar la fecha de publicación de otra cita o por una razón especial (como la indicada).		
El presente informe ha sido realizado ☒ para todas las reivindicaciones ☐ para las reivindicaciones nº:		
Fecha de realización del informe 28.11.2022	Examinador J. M. Vazquez Burgos	Página 1/2

CLASIFICACIÓN OBJETO DE LA SOLICITUD

G06F16/332 (2019.01)

G06F40/30 (2020.01)

G16H10/20 (2018.01)

G10L15/26 (2006.01)

G10L13/00 (2006.01)

H04L51/02 (2022.01)

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G06F, G16H, G10L, H04L

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, INTERNET