ENTREGABLE
PROYECTOS—
2023-2024

Investigación en sistemas de conversión de lenguaje natural para la programación de robots en tareas colaborativas
"NATURBOT"

Entregable: E4.1. Sistema de interpretación desarrollado

Programa: Proyectos de I+D en colaboración con empresas
Número de proyecto: 22300046
Expediente: IMDEEA/2023/13
Duración: 16 meses
Coordinado en AIDIMME por: José Luís Sánchez

GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea

AIDIMME
INSTITUTO TECNOLÓGICO

ÍNDICE

1 Introducción y objetivos del entregable _____ 1

1.1 Objeto del entregable _____ 1

1.2 Objetivos del proyecto con los que se relaciona _____ 1

1.3 Paquete de trabajo en el que está encuadrado y principales tareas a desarrollar _ 1

1.4 Resumen del trabajo realizado _____ 1

2 Actividades realizadas _____ 3

2.1 Tarea 4.1 Definición de ontologías en la interacción humano-robot _____ 3

2.1.1 Introducción al concepto de ontología _____ 3
2.1.2 Definición de los elementos básicos de la ontología a desarrollar _____ 4
2.1.3 Gestión de conceptos equivalentes (tesauros) _____ 5
2.1.4 Desarrollo de la ontología de ámbito específico de aplicación _____ 6

2.2 Tarea 4.2 Investigación sobre sistemas de interpretación de lenguaje natural ___ 8

2.3 Tarea 4.3 Generación de algoritmos y desarrollo de la aplicación _____ 11

2.3.1 Introducción a Amazon Lex _____ 11
2.3.2 Definición del dominio de trabajo _____ 12
2.3.3 Definición inicial de las peticiones del usuario al sistema _____ 13
2.3.4 Implementación de la plantilla en Amazon Lex _____ 15
2.3.5 Testeo en la propia plataforma AWS _____ 23
2.3.6 Desarrollo del modelo _____ 23
2.3.7 Desarrollo interfaz de usuario _____ 27

2.4 Tarea 4.4 Desarrollo de pruebas _____ 28

3 Resultados obtenidos _____ 29

4 Conclusiones _____ 32

5 Anexos y bibliografía _____ 33

1 Introducción y objetivos del entregable

1.1 Objeto del entregable

El presente entregable recoge toda la documentación asociada al sistema de interpretación del lenguaje natural basado en órdenes verbales desarrollado dentro del proyecto.

Este sistema es el encargado de interpretar las órdenes verbales emitidas por un usuario en lenguaje natural (por voz) y asignarle una instrucción de trabajo específica que posteriormente será ejecutada por un robot.

El desarrollo de este sistema es un primer resultado del proyecto que permite avanzar en los métodos de interacción humano-robot, y genera el primer componente de un sistema final que debe de integrar, además, un sistema de reconocimiento del entorno (paquete de trabajo PT5) y un sistema de comunicación de comandos en un lenguaje interpretable por un robot (paquete de trabajo PT6).

1.2 Objetivos del proyecto con los que se relaciona

Este documento se relaciona con dos de los objetivos específicos identificados en la memoria de solicitud del proyecto:

- Investigación en sistemas de interpretación del lenguaje natural, para descomponer expresiones habitualmente utilizadas en tareas colaborativas, en órdenes elementales.
- Desarrollo de un subsistema basado en IA para la realización de dicha transformación.

1.3 Paquete de trabajo en el que está encuadrado y principales tareas a desarrollar

El entregable se encuadra en el paquete de trabajo PT4 (Investigación en sistemas de interpretación de lenguaje natural), donde se llevan a cabo las siguientes tareas:

- Tarea 4.1 Definición de ontologías en la interacción humano-robot
- Tarea 4.2 Investigación sobre sistemas de interpretación de lenguaje natural
- Tarea 4.3 Generación de algoritmos y desarrollo de la aplicación
- Tarea 4.4 Desarrollo de pruebas

1.4 Resumen del trabajo realizado

En primer lugar, se han estudiado las posibles relaciones humano-robot, con especial énfasis en aquellas tareas colaborativas en las que pudiera tener cabida la comunicación por voz mediante lenguaje natural. Se ha definido una ontología que representa el conocimiento y los términos asociados a estas relaciones.

En segundo lugar, se ha llevado a cabo una revisión de sistemas de interpretación del

AIDIMME INSTITUTO TECNOLÓGICO · GENERALITAT VALENCIANA · IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN · Financiado por la Unión Europea

lenguaje natural basados en voz identificando aquellos que mayor aplicabilidad tienen en el proyecto.

A continuación, se ha desarrollado un bot conversacional basado en la tecnología de AWS (Amazon Lex) que interpreta las solicitudes que puede realizar un usuario para un escenario de trabajo en la que se solicita la ayuda de un robot para que le acerque o devuelva un conjunto de piezas con colores y tamaños definidos.

Se ha desarrollado también una interfaz de usuario que se complementa con el bot que interpreta las instrucciones del usuario.

Finalmente, se han llevado a cabo pruebas preliminares de funcionamiento con diferentes usuarios con el objetivo de validar el funcionamiento del sistema.

AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea

2 Actividades realizadas

2.1 Tarea 4.1 Definición de ontologías en la interacción humano-robot

2.1.1 Introducción al concepto de ontología

El principal objetivo de esta tarea es la propuesta y definición de una ontología que represente el dominio abordado por el escenario de aplicación propuesto en el proyecto, el cual esencialmente consiste en la colaboración entre humano y robot mediante la asignación de tareas sencillas de envío y recogida de piezas a/desde el puesto del trabajador.

Según (Neches et al., 1991) una ontología define los conceptos básicos y las relaciones que comprenden el vocabulario de un ámbito en particular, así como las reglas para combinar estos términos y relaciones para definir extensiones del vocabulario.

Las ontologías se pueden clasificar (Guarino, 1998) en base a su nivel de dependencia respecto a una tarea en particular, de la siguiente manera:

- Ontologías de alto nivel: describen conceptos muy genéricos (p.ej., espacio, tiempo, materia, objetos, eventos, acciones, etc.) que son independientes de algún problema o domino concreto, y suelen utilizarse en grandes comunidades de usuarios.
- Ontologías de dominio y de tarea: describen el vocabulario relacionado con un dominio concreto (p.ej., medicina, automoción) o una tarea genérica (p.ej., diagnóstico, ventas), mediante la especialización de términos introducidos en una ontología de alto nivel.
- Ontologías de aplicación: describen conceptos dependiendo de un dominio o una tarea en particular, las cuales suelen ser especializaciones de ambas ontologías relacionadas. A menudo estos conceptos corresponden a roles ejecutados por entidades de dominio realizando una actividad en concreto (unidad de recambio, pieza).

Cabe también diferenciar entre una ontología de aplicación y una base de conocimiento, pues la diferencia radica en el objetivo de la ontología, que se corresponde con una base de conocimiento específica, describiendo hechos que se consideran válidos por una comunidad de usuarios en base al significado acordado del vocabulario que se utilice. Una base de conocimiento genérica puede describir hechos relacionados con un estado concreto, y dentro de esta base se pueden distinguir dos elementos: la ontología (que contiene información que es independiente del estado), y una base de conocimiento "principal" (que contiene información dependiente del estado) (Guarino, 1998).

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

2.1.2 Definición de los elementos básicos de la ontología a desarrollar

En el ámbito de aplicación de proyecto, definimos una ontología específica que cubre los procesos de entrega y devolución de piezas a un operario por parte de un robot colaborativo, el cual recibe órdenes por voz para ejecutar dichas tareas siguiendo una estructura de lenguaje apropiada.

Para ello, es necesario identificar los recursos que forman parte del dominio propuesto de aplicación, las cuales son principalmente: Entidades, Relaciones, y Propiedades. A continuación, se muestran los elementos identificados para cada categoría:

Entidades

- Operario
- Robot colaborativo
- Pieza
- Localizaciones
    a. Mesa de trabajo (Puesto del Operario)
    b. Mesa de almacenamiento de piezas

Relaciones

Estas básicamente serían las acciones (intenciones) que se pueden realizar en el entorno: el Operario puede solicitar al Robot las siguientes acciones:

- Recibir una Pieza: El Robot Lleva una Pieza a la Mesa de trabajo (puesto del Operario)
- Devolver una pieza: El Robot Devuelve una Pieza a la Mesa de almacenamiento

Además de las anteriores, cabe tener en cuenta relaciones inherentes al escenario de aplicación que es recomendable representar en la ontología, p.ej.:
- El Operario “solicita” al Robot la realización de la acción correspondiente
- El Robot “ejecuta” la orden solicitada por el Operario
- El Robot puede “portar” la Pieza solicitada en un momento dado
- Las Piezas “se ubican” en alguna de las Localizaciones destinadas para ello
- Etc.

Propiedades

De las Entidades identificadas, a priori la única que contiene propiedades relevantes es la Pieza:

- Color (p.ej., amarilla, roja, azul)
- Tamaño (p.ej., pequeña (4 pines), grande (6 pines))

AIDIMME INSTITUTO TECNOLÓGICO · GENERALITAT VALENCIANA · IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN · Financiado por la Unión Europea

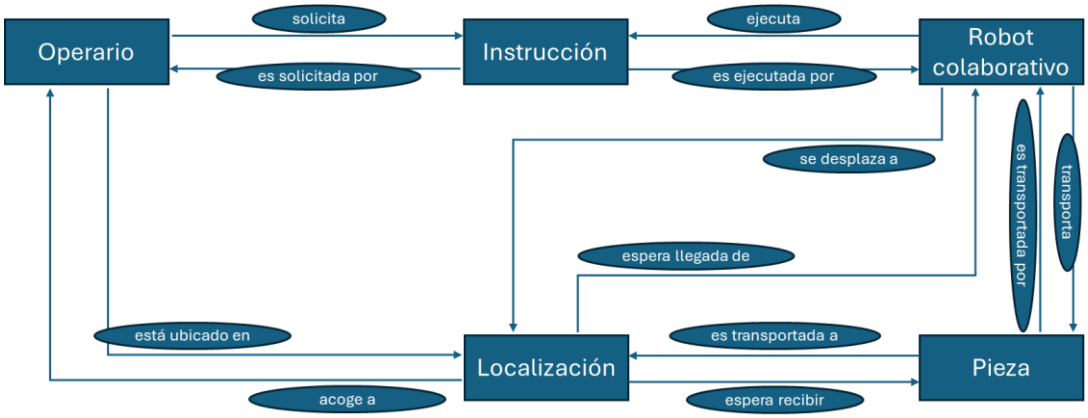


Figura 1: diagrama de relaciones más relevantes entre entidades de la ontología desarrollada
Fuente: elaboración propia

Cabe destacar que las Propiedades *Color* y *Tamaño* han sido representadas como Entidades en la ontología desarrollada para poder incluirse como Rango (posibles valores) de otras entidades (p.ej., Pieza).

2.1.3 Gestión de conceptos equivalentes (tesauros)

Uno de los aspectos más importantes del escenario de aplicación para tener en cuenta en el desarrollo de la ontología de comunicación humano-robot es la gestión de los términos sinónimos, pues el robot colaborativo debe ser capaz de interpretar determinados que son susceptibles de ser expresados utilizando distintos términos. Para ello, se ha definido una lista de términos equivalentes que puede ser extendida según necesidades específicas. Esto afecta a Entidades, Relaciones, y Propiedades, como se muestra a continuación.

<u>Entidades</u>

- Pieza: Bloque, Ladrillo
- Mesa de trabajo: Puesto, Sitio del Operario
- Mesa de almacenamiento: Almacén, Mesa de piezas

<u>Relaciones</u>

- Acercar Pieza: Trae, Dame, Entrégame, Acerca, Quiero utilizar, Quiero usar, Necesito, Voy a usar, Tráeme, Acércame, Me hace falta
- Devolver Pieza: Toma, Devuelve, Retira, Aparta, Retorna, No necesito, Ya no necesito, No me hace falta, La puedes devolver

<u>Propiedades</u>

(En particular, de la Entidad Pieza)

- Color
    a. Verde: color verde, la verde, la de color verde

   b. Amarilla: Amarillo, Color amarillo, La amarilla, La de color amarillo, dorado/a
   c. Azul: Azulado/a, Color azul, la azul, la de color azul
   d. Roja: La roja, La de color rojo
- Tamaño
   a. Pequeña: Pequeño, Más pequeño/a, Corto/a, Fino/a, Estrecho/a, Chico/a, Tamaño pequeño, La pequeña, cuatro pines
   b. Grande: Alargado/a, más grande, alto/a, ancho/a, enorme, voluminoso/a, grueso/a, tamaño grande, la grande, seis pines

En la definición de la ontología se utilizan identificadores únicos globales (URI por sus siglas en inglés) para referirse a los recursos en ella representados de manera inequívoca. Las URIs que puedan estar asignadas a los conceptos de un tesauro permiten referirse a estos conceptos inequívocamente desde cualquier contexto. De este modo, el nombre preferente de un término vendría a ser una propiedad preferente asignada al concepto, y cada nombre no preferente sería una propiedad alternativa para ese concepto. Para representar estas propiedades utilizamos las propiedades *skos:prefLabel* y *skos:altLabel* para cada caso, que permiten asignar una etiquetas léxica preferente y etiquetas léxicas alternativas a un mismo recurso (Vilches-Blázquez, 2009).

2.1.4 Desarrollo de la ontología de ámbito específico de aplicación

La ontología para el entorno de aplicación del escenario ha sido desarrollada mediante el software Protégé en formato Web Ontology Language (OWL). El fichero generado contiene las entidades, propiedades y relaciones necesarias que conforman el escenario propuesto.

A continuación, se muestran algunas imágenes que ilustran algunos aspectos de la ontología desarrollada:

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

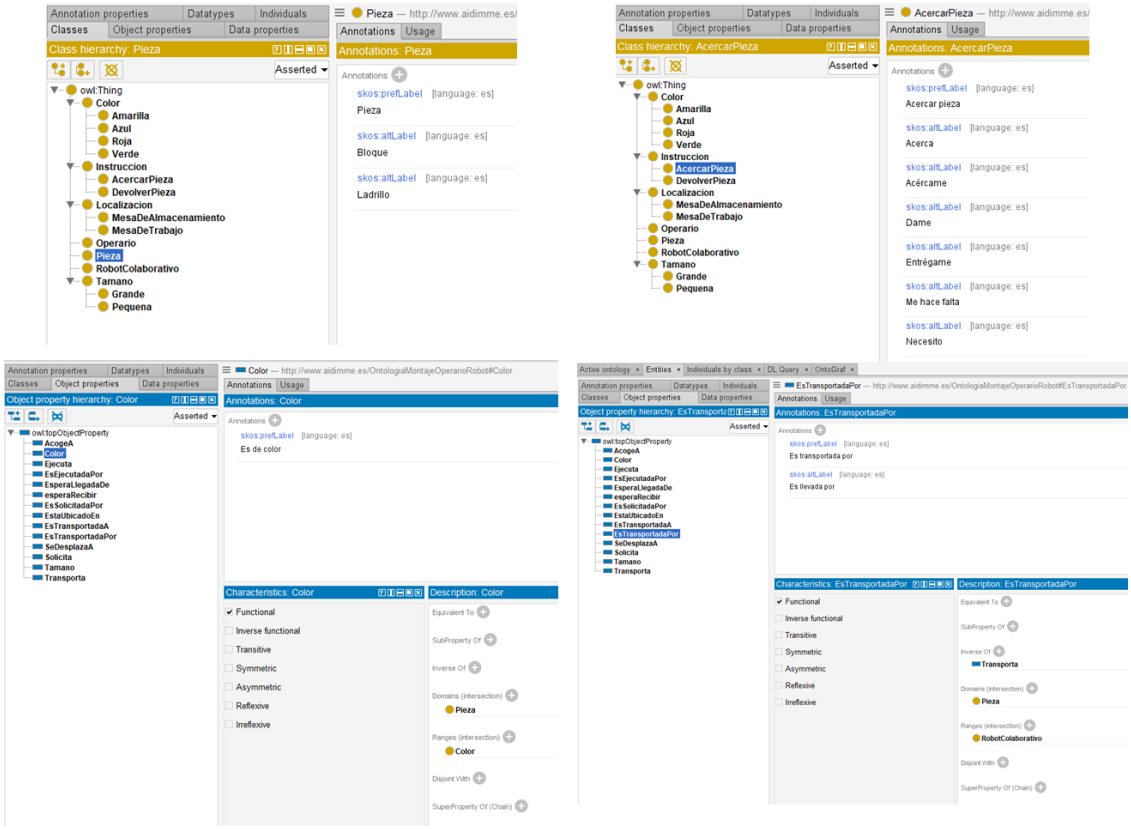


Figura 2: extracto de diferentes elementos definidos en la ontología para el entorno de aplicación
Fuente: elaboración propia

En la parte superior de la imagen se muestran ejemplos de Entidades y la gestión de sus vocablos sinónimos mediante el uso de la anotación *skos:altLabel*. En la parte inferior se muestran ejemplos de relaciones representadas como Propiedades, como es el color de una pieza, o el transporte de piezas por parte del robot colaborativo.

La imagen a continuación muestra el mapa de relaciones entre las distintas Entidades definidas en la ontología.

AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea

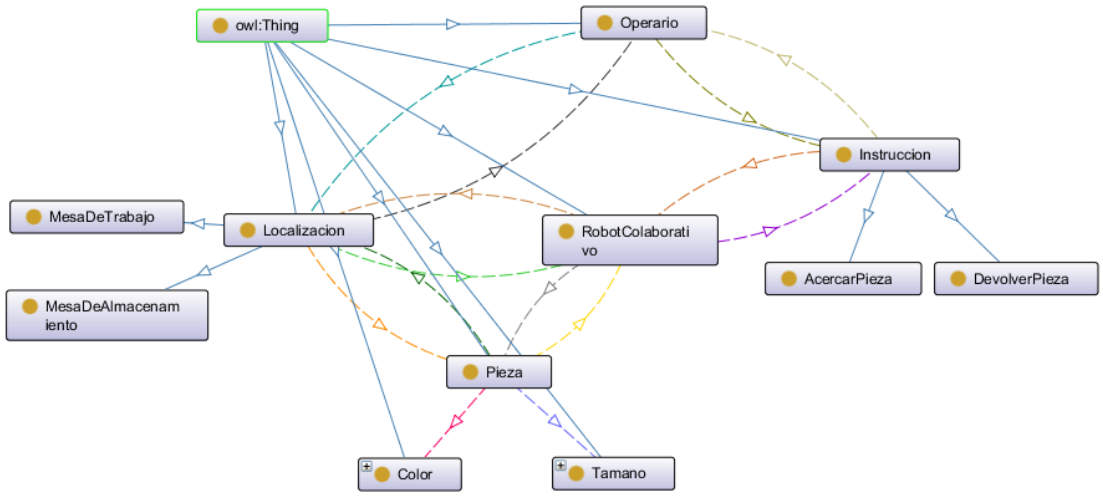


Figura 3: mapa de relaciones entre las entidades definidas en la ontología para el caso de uso
Fuente: elaboración propia

Como puede observarse en la figura, el mapa de relaciones entre Conceptos generado es relativamente sencillo, pues el dominio de aplicación considera fundamentalmente las características del caso de uso, focalizado en el envío y retirada de piezas a/desde el puesto del trabajo del trabajador mediante el robot colaborativo.

Sin embargo, el diseño de la ontología permite su ampliación mediante la inserción de nuevas Entidades, Propiedades y Relaciones según actualizaciones del escenario o su aplicación en entornos de distintas características.

2.2 Tarea 4.2 Investigación sobre sistemas de interpretación de lenguaje natural

Existen diferentes sistemas de reconocimiento de voz, que permiten trasladar lo que habla una persona (audio) a una transcripción en forma de texto, lista para poder ser interpretada.

Por un lado, se disponen de modelos de código abierto que pueden ser integrados en una aplicación como la que se pretende con NATURBOT, como, por ejemplo:

- CMU Sphinx: Un motor de código abierto con alta precisión y soporte para múltiples idiomas.
- Julius: Otro motor de código abierto con soporte para una amplia gama de características acústicas y de lenguaje.
- Vosk: Un motor de código abierto ligero y rápido, ideal para dispositivos móviles y aplicaciones con recursos limitados.

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

También existen servicios en la nube de distintos proveedores, que, a cambio de un pago por uso, proporcionan el servicio de transcripción:

- Google Cloud Speech-to-Text: Una API de pago que ofrece alta precisión, escalabilidad y soporte para múltiples idiomas y formatos de audio.
- Amazon Transcribe Speech-to-Text: Otra API de pago con características similares a la de Google Cloud Speech-to-Text.
- Microsoft Azure Speech Services: Una API de pago que ofrece reconocimiento de voz, síntesis de voz y traducción automática.
- IBM Watson Speech to Text: Una API de pago con soporte para modelos acústicos personalizados y herramientas de análisis del habla.

También se han identificado algunas soluciones preconstruidas:

- Speechnotes: Una aplicación web gratuita que ofrece dictado de voz en tiempo real con transcripción y puntuación automáticas.
- Dictation.io: Otra aplicación web gratuita similar a Speechnotes.
- Rev.ai: Una plataforma de transcripción de audio y vídeo que ofrece servicios profesionales de transcripción manual y automática.
- Trint: Otra plataforma de transcripción de audio y vídeo similar a Rev.ai.

Todos estos sistemas se basan en la transcripción del audio a texto, pero no en su interpretación, que es lo que se busca con NATURBOT.

Un sistema de interpretación del lenguaje natural por voz va más allá. Se trata una tecnología diseñada para permitir la interacción entre humanos y computadoras a través del habla. Utilizando algoritmos avanzados de procesamiento del lenguaje natural (PLN) y reconocimiento de voz, estos sistemas son capaces de comprender y procesar los comandos y consultas emitidas por los usuarios mediante su voz. Sus principales características incluyen la capacidad de transcribir la voz a texto con alta precisión, entender el significado y la intención detrás de las palabras pronunciadas, y tomar acciones en función de esa comprensión. Además, algunos sistemas también pueden generar respuestas habladas o texto a partir de las solicitudes del usuario, lo que permite una interacción bidireccional natural.

Estos sistemas son actualmente utilizados en algunas aplicaciones domésticas o de oficinas, como asistentes virtuales en teléfonos inteligentes, o interfaces de usuario en dispositivos domésticos y aplicaciones empresariales. Su capacidad para comprender el lenguaje humano en tiempo real y actuar según las instrucciones proporcionadas por los usuarios los hace útiles en entornos donde la entrada de texto puede ser difícil o inconveniente, como en la conducción o en entornos industriales donde el usuario requiere tener las manos libres para trabajar.

AIDIMME INSTITUTO TECNOLÓGICO · GENERALITAT VALENCIANA · IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN · Financiado por la Unión Europea

En el presente proyecto se aplica esta tecnología para interpretar las acciones que un usuario desea que un robot ejecute en tareas colaborativas en un entorno industrial. El reto consiste en lograr la aplicación de una tecnología del ámbito doméstico o de oficinas a un ámbito industrial donde se requiere mayor precisión y velocidad de respuesta, en un entorno con ruido de maquinaria u otros operarios.

En los últimos diez años, han surgido diferentes sistemas encaminados a ofrecer servicios de interpretación de lenguaje natural por voz. Se ha llevado a cabo una revisión y análisis de este tipo de sistemas y la siguiente tabla lo resume especificando algunas ventajas e inconvenientes detectados.

Nombre del Sistema	Empresa	Año de Lanzamiento	Ventajas	Inconvenientes
AWS Lex	Amazon	2016	- Integración con otros servicios de AWS como Lambda y Polly. - Fácil de usar con la consola de AWS. - Escalabilidad y fiabilidad debido a la infraestructura de AWS.	- La personalización puede requerir habilidades de desarrollo. - Algunas características avanzadas pueden ser limitadas en comparación con otros sistemas.
Google Dialogflow	Google	2016	- Potente capacidad de procesamiento del lenguaje natural. - Integración con Google Cloud Platform y otros servicios de Google. - Soporte para múltiples canales de entrada y salida.	- La configuración inicial puede ser compleja para usuarios no técnicos. - Algunas características avanzadas pueden requerir conocimientos de desarrollo avanzado.
IBM Watson	IBM	2013	- Amplia gama de capacidades de procesamiento del lenguaje natural, incluyendo análisis de sentimientos y extracción de entidades. - Integración con servicios de IBM Cloud. - Personalización y entrenamiento del modelo avanzado.	- El precio por uso puede ser elevado para ciertos casos. - La curva de aprendizaje puede ser elevada para usuarios no técnicos.
Microsoft Azure Language Understanding	Microsoft	2017	- Integración con servicios de Microsoft Azure, como Azure Bot Service y Azure Functions. - Herramientas de entrenamiento y evaluación fáciles de usar. - Soporte para varios idiomas y canales de entrada.	- Limitaciones en la personalización avanzada del modelo. - Algunas características pueden requerir conocimientos técnicos elevados para la configuración.

Salvo el sistema de IBM que presenta un coste de uso bastante elevado, y funcionalidades ligadas al análisis de sentimientos que no se prevé usar en el proyecto,

AIDIMME INSTITUTO TECNOLÓGICO · GENERALITAT VALENCIANA · IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN · Financiado por la Unión Europea

el resto de los sistemas (AWS Lex, Google Dialogflow y Microsoft Azure Language Understanding) presentan características similares y su potencial aplicabilidad en el proyecto es alta.

El desarrollo final de la aplicación se va a llevar a cabo en AWS Lex debido a su entorno de desarrollo amigable y a facilidad de escalado de los desarrollos que se realicen. Todos estos sistemas son sistemas acotados y guiados, suficiente para el ámbito de este proyecto.

2.3 Tarea 4.3 Generación de algoritmos y desarrollo de la aplicación

2.3.1 Introducción a Amazon Lex

Amazon Lex es un servicio de AWS (Amazon Web Services) que permite la creación de interfaces de conversación de voz y texto, conocidas como chatbots o bots de voz. Estos chatbots pueden interactuar con los usuarios de manera natural, entendiendo preguntas y comandos en lenguaje natural y respondiendo de manera automatizada.

El funcionamiento básico de Amazon Lex es el siguiente:

- Definición del bot: El primer paso es definir un bot en Amazon Lex. Esto implica diseñar las intenciones (intentions) que el bot debe entender, configurar las expresiones que los usuarios pueden utilizar para comunicarse con el bot y definir las respuestas o acciones que el bot debe tomar en función de las intenciones detectadas.
- Entrenamiento del modelo de lenguaje: Una vez que se ha definido el bot, se debe entrenar el modelo de lenguaje. Amazon Lex utiliza aprendizaje automático para comprender el lenguaje natural de los usuarios. Durante el entrenamiento, el servicio analiza las expresiones de los usuarios y las relaciona con las intenciones y entidades definidas. La ontología definida en la Tarea 4.1 servirá de guía para explicitar las “intenciones” y el resto de las entidades que se requieren para el entrenamiento.
- Implementación del bot: Una vez que el modelo de lenguaje está entrenado, se puede implementar el bot. Esto implica configurar cómo el bot interactuará con los usuarios, ya sea a través de texto o voz.
- Integración con canales de comunicación: Amazon Lex puede integrarse con diversos canales de comunicación, como aplicaciones web, aplicaciones móviles, mensajería instantánea, dispositivos de voz, entre otros. Se puede configurar estas integraciones para que los usuarios puedan interactuar con el bot a través de los canales que prefieran.
- Interacción con los usuarios: Cuando los usuarios interactúan con el bot a través de un canal de comunicación, Amazon Lex procesa sus consultas y utiliza el modelo de

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

lenguaje entrenado para comprender sus intenciones y entidades. Luego, determina la mejor respuesta o acción a tomar.

- Acciones y respuestas del bot: Basado en lo que el bot ha entendido de la interacción del usuario, puede realizar diversas acciones, como consultar bases de datos, realizar cálculos, acceder a servicios web o proporcionar respuestas predefinidas. Puede responder al usuario en forma de texto o voz, según el canal de comunicación.
- Gestión de diálogo y contexto: Amazon Lex tiene capacidades para gestionar conversaciones continuas y recordar el contexto de la conversación. Esto permite que los usuarios realicen preguntas de seguimiento o tengan conversaciones más largas sin tener que repetir información.

2.3.2 Definición del dominio de trabajo

Se plantea un escenario de trabajo inicial donde el usuario quiere realizar un montaje con diferentes tipos de piezas. El usuario dará instrucciones verbales en lenguaje natural al robot indicando que piezas desea utilizar, y el robot debe ejecutar las acciones que le han sido solicitadas.

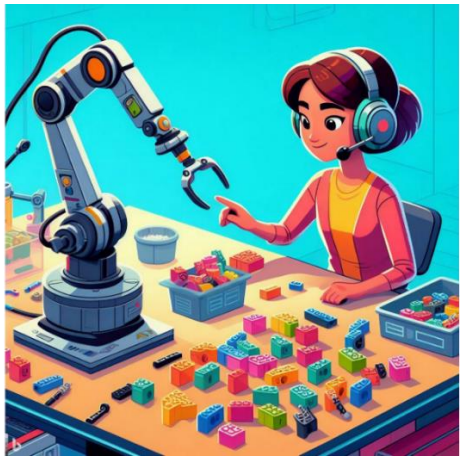


Figura 4: Recreación del posible escenario de trabajo
Fuente: elaboración propia, mediante Bing Image Creator

El SISTEMA VOZ debe de interpretar el lenguaje natural empleado por el usuario, entender que acción quiere llevar a cabo el usuario y con qué tipo de pieza. Además, deberá de comunicar la información adecuada para que se activen los sistemas de reconocimiento del entorno.

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

Figura 5: esquema de funcionamiento general de la solución
Fuente: elaboración propia

Tal y como se ha definido en la ontología de la Tarea 4.1, en este entorno de trabajo, el usuario puede desarrollar 2 tipos de acciones (intentions), utilizando 6 tipos diferentes de piezas.
Acciones a realizar por el robot:

- Coger una pieza de lego de la mesa y colocarla en la zona de la mesa próxima al usuario
- Devolver una pieza que ya no quiere el usuario a la zona de la mesa donde están el resto de las piezas.

Piezas a manipular:

- Piezas amarillas pequeñas o grandes
- Piezas rojas pequeñas o grandes
- Piezas azules pequeñas o grandes

2.3.3 Definición inicial de las peticiones del usuario al sistema

Para la utilización del sistema AWS Lex cada posible petición del usuario al sistema debe de basarse en el reconocimiento de "intenciones". Estas "intenciones" (intentions, en el entorno de desarrollo de AWS Lex) son representaciones de las acciones o propósitos que un usuario puede tener al interactuar con un chatbot o una interfaz de voz. Cada intención define qué tipo de solicitud o pregunta está realizando el usuario.

Cada "intention" en Amazon Lex debe tener información asociada que defina su propósito, cómo se debe manejar y cómo interactuar con los usuarios. La información asociada a una intención puede incluir los siguientes aspectos:

- Nombre de la intención: Un nombre descriptivo que identifica de manera única la intención. Este nombre se utiliza para referirse a la intención en la configuración y programación.

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

- Ejemplos o expresiones: Una lista de ejemplos de expresiones que los usuarios podrían utilizar para activar la intención. Estos ejemplos sirven para entrenar el modelo de lenguaje y permiten a Amazon Lex comprender cuando un usuario está expresando esa intención.
- Respuestas o acciones: La lógica que define lo que el chatbot debe hacer cuando se activa esta intención. Esto puede incluir la ejecución de una función, la búsqueda en una base de datos, el acceso a servicios web, el cálculo de una respuesta o cualquier otra acción específica para satisfacer la intención del usuario.
- Parámetros o entidades: Si la intención requiere información adicional del usuario para completarse, se debe definir los parámetros o entidades asociados a la intención.
- Preguntas de clarificación: Se pueden definir preguntas de clarificación que el chatbot realizará si no entiende completamente la intención del usuario o necesita más información. Esto permite aclarar la solicitud del usuario y garantizar una interacción más precisa.

A continuación, se muestra cómo se definen dos “intention” básicas que se trabajarán durante las pruebas del sistema:

- Que el robot entregue una pieza al operario
- Que el robot devuelva una pieza a su puesto de trabajo original

A continuación, se desarrolla la información asociada a cada una de estas dos “intentions”.

2.3.3.1 Coger una pieza y colocarla en la zona de la mesa próxima al usuario

Intention: Recibir una pieza que necesita el usuario
Entidades: color (amarillo, rojo, azul), tamaño (pequeña, grande)
Locuciones:

- Para “dar”: Usar, utilizar, dar, necesitar, hacer falta
- Para “color”: amarillo, rojo, azul
- Para “Tamaño”: pequeña, grande, corta, larga, alargada, 4 pines, 8 pines,

Ejemplos:

- Necesito que me des la pieza grande roja
- Me hace falta la amarilla más pequeña
- Voy a utilizar la pieza azul de 8 pines

Respuestas: ¿Quieres que te de la pieza <<color>> de <<tamaño>>?

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

2.3.3.2 Devolver una pieza que ya no quiere el usuario a la zona de la mesa donde están el resto de las piezas.

Intention: Devolver una pieza que ya no necesita el usuario
Entidades: color (amarillo, verde, azul), tamaño (pequeña, grande)
Locuciones:

- Para "devolver": Devolver, retornar, no necesitar, no hacer falta
- Para "color": amarillo, roja, azul
- Para "Tamaño": pequeña, grande, corta, larga, alargada, 4 pines, 8 pines,

Ejemplos:

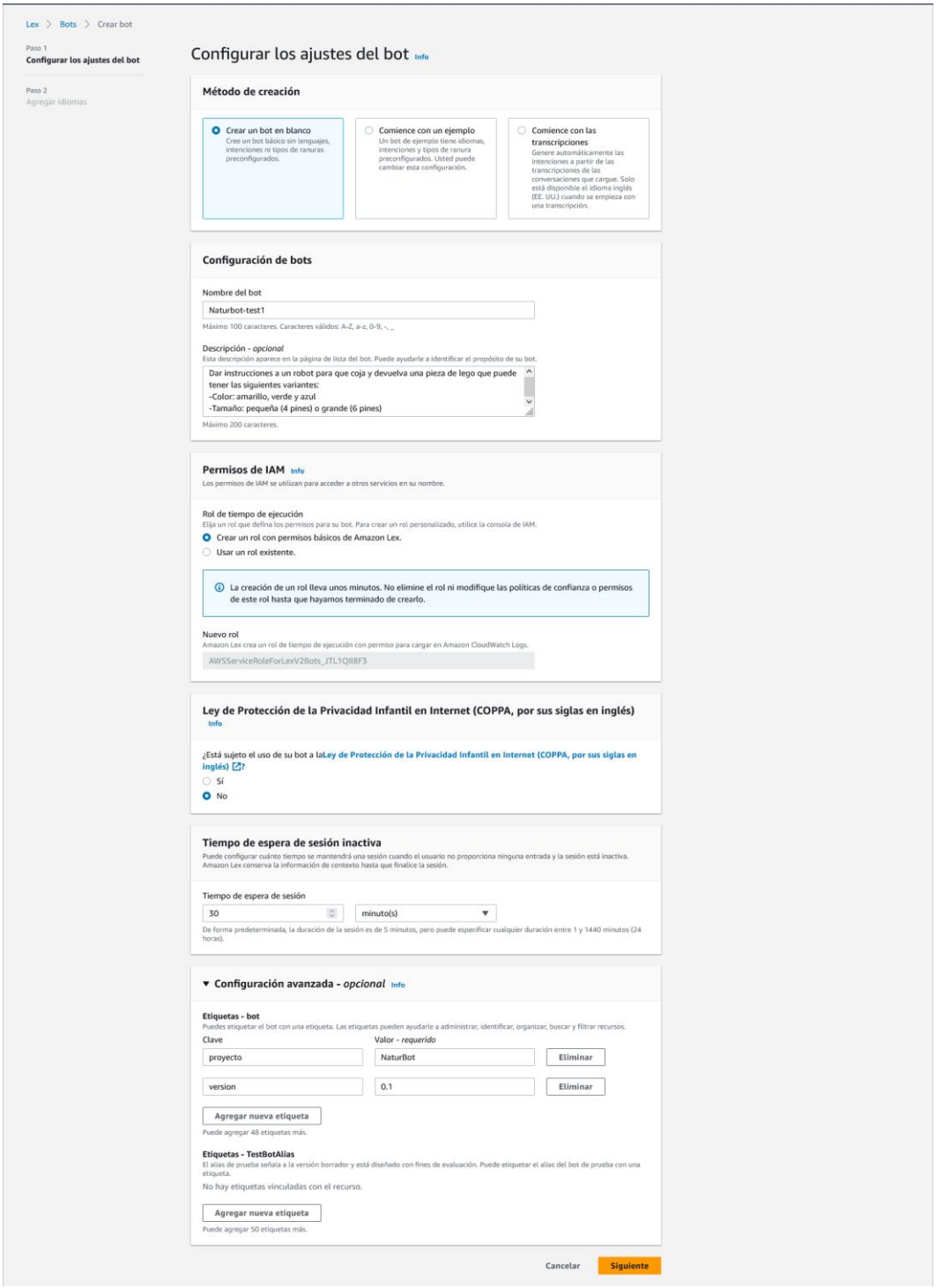
- Ya no necesito la pieza grande roja
- No me hace falta la amarilla más pequeña
- Voy a devolver la pieza azul de 8 pines

Respuestas: ¿Quieres que devuelva a la mesa de trabajo la pieza <<color>> de <<tamaño>>?

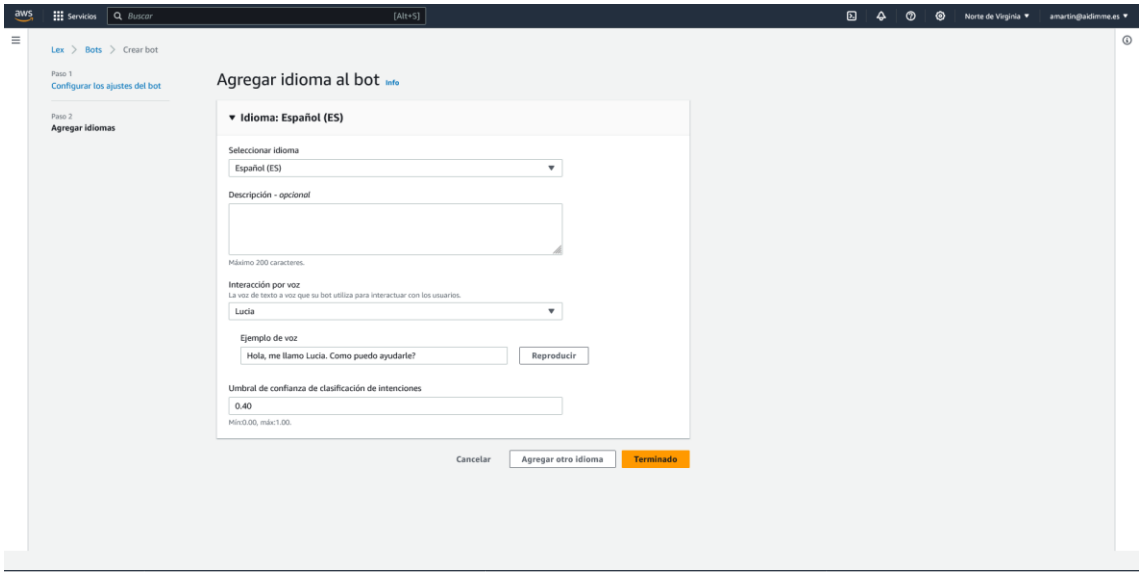
2.3.4 Implementación de la plantilla en Amazon Lex

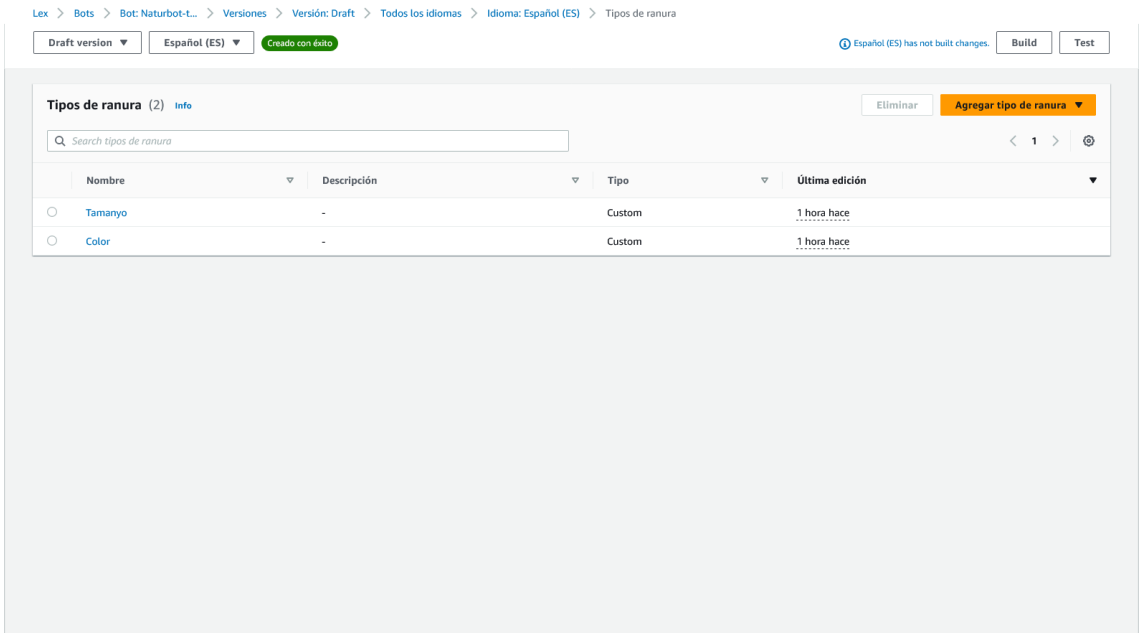
A continuación, se muestran diferentes capturas de pantalla del proceso de configuración de un bot para trabajar en un entorno de trabajo como el descrito, donde el operario puede pedir al robot realizar acciones de entregar o devolver una pieza.

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

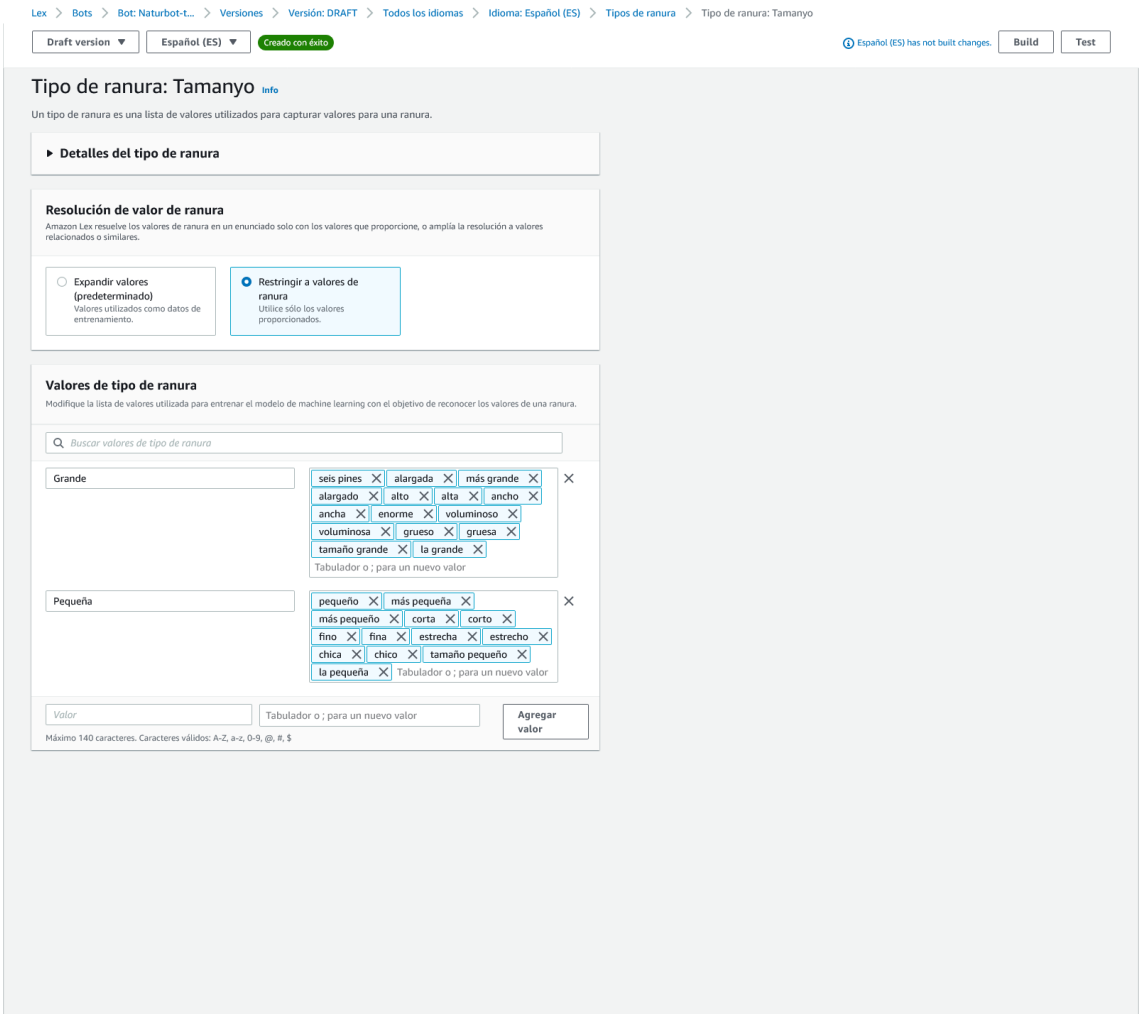


AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea

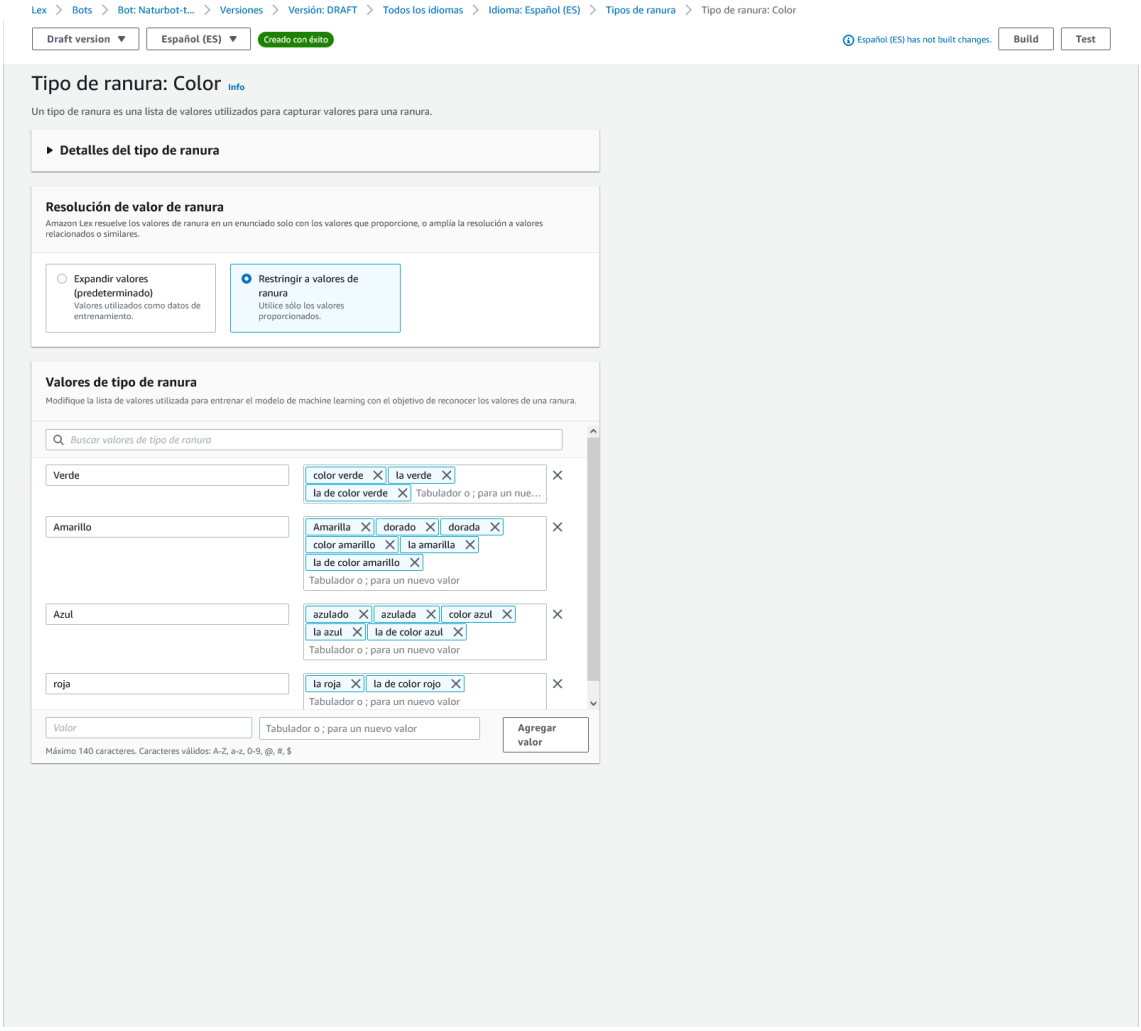




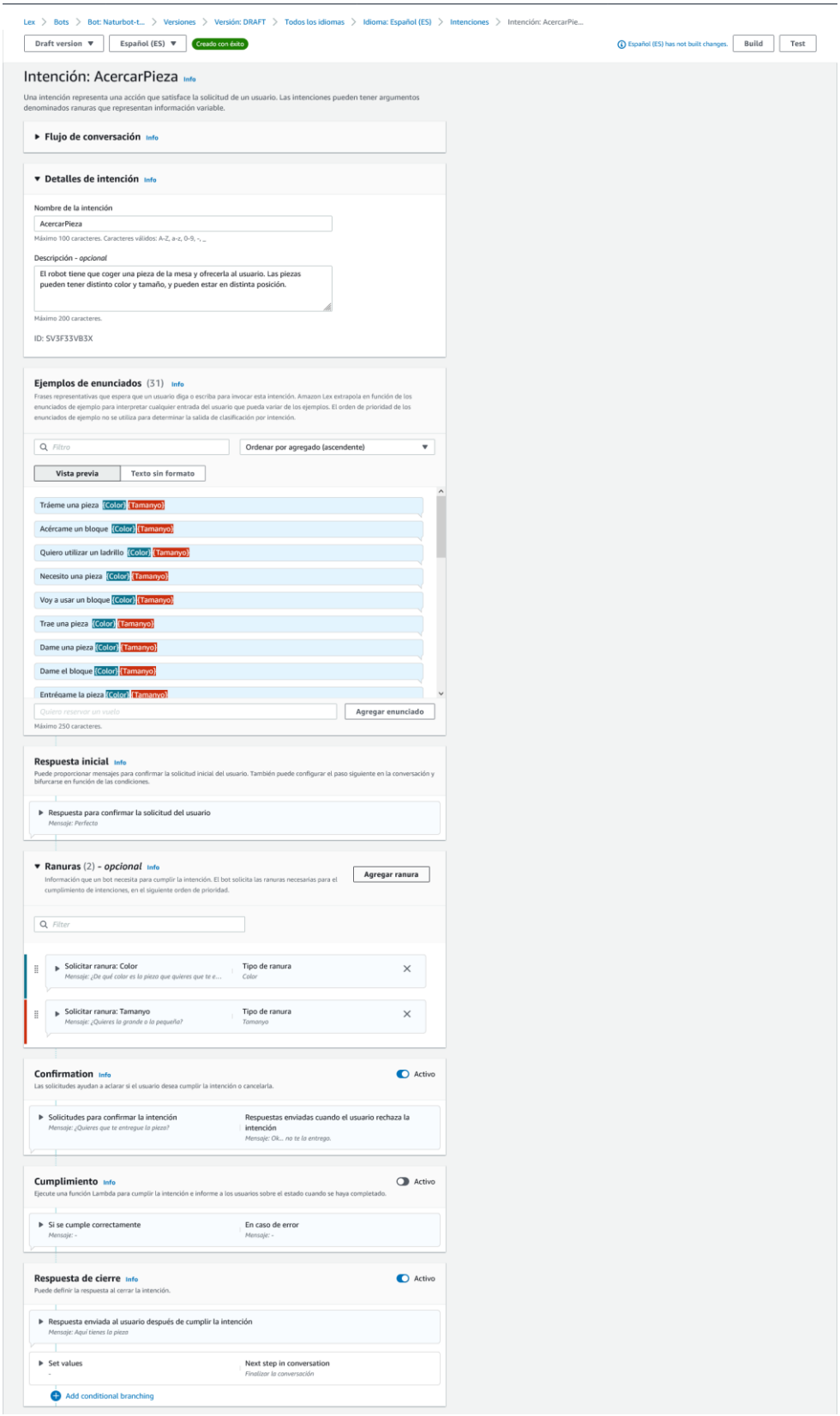
AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea



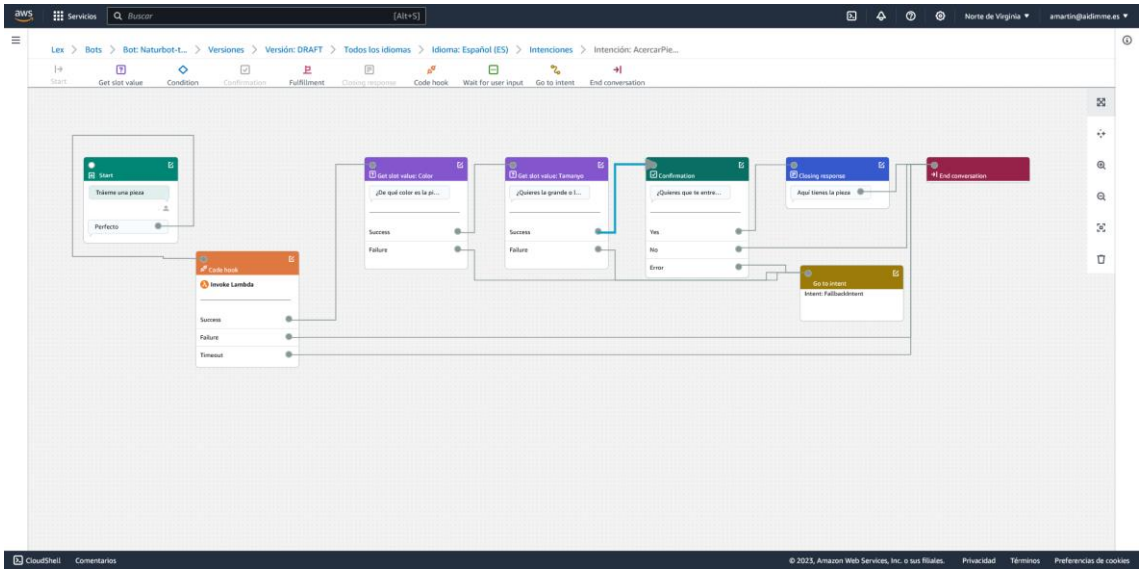
AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea



AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea



AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea



AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea

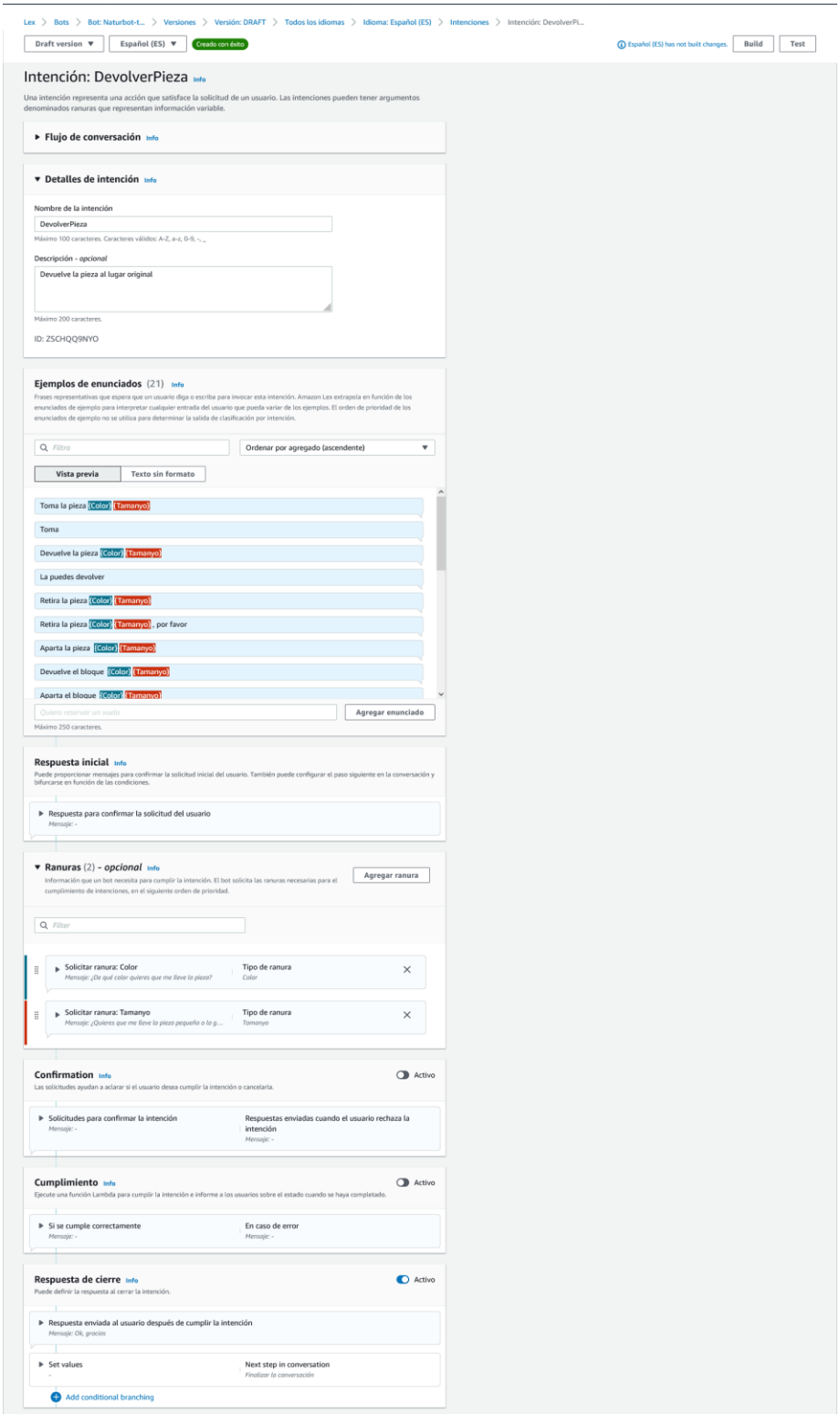


Figura 6: capturas de pantalla de implementación del bot
Fuente: elaboración propia, en plataforma de AWS

AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea

2.3.5 Testeo en la propia plataforma AWS

A continuación, se muestra una de las pruebas de ejemplo realizadas con el bot conversacional desarrollado para el caso de uso de la manipulación de las 6 piezas, en la propia plataforma AWS.

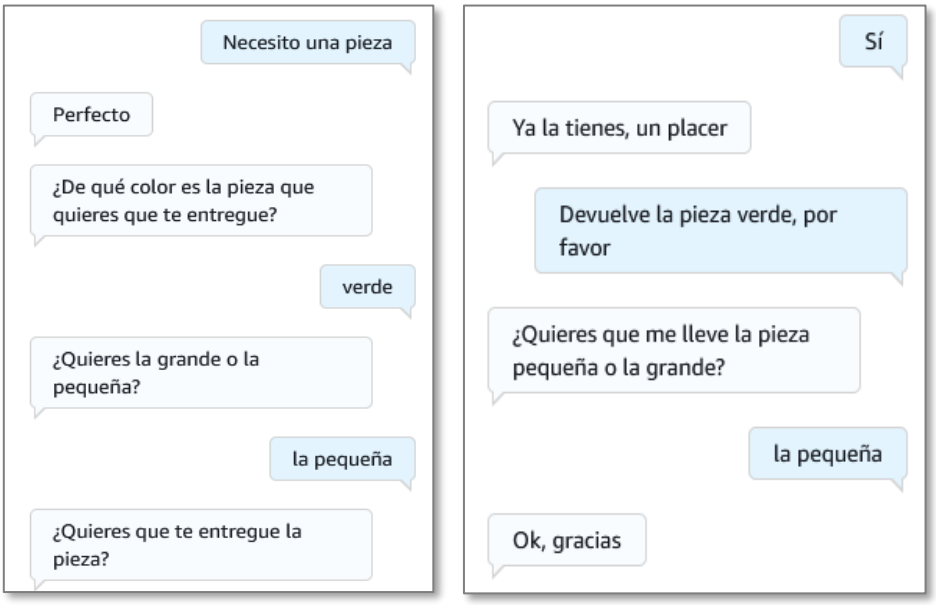


Figura 7: ejemplo de pruebas iniciales en plataforma AWS
Fuente: elaboración propia, en plataforma de AWS

2.3.6 Desarrollo del modelo

El desarrollo de los distintos elementos que componen el sistema se basa en el esquema siguiente:

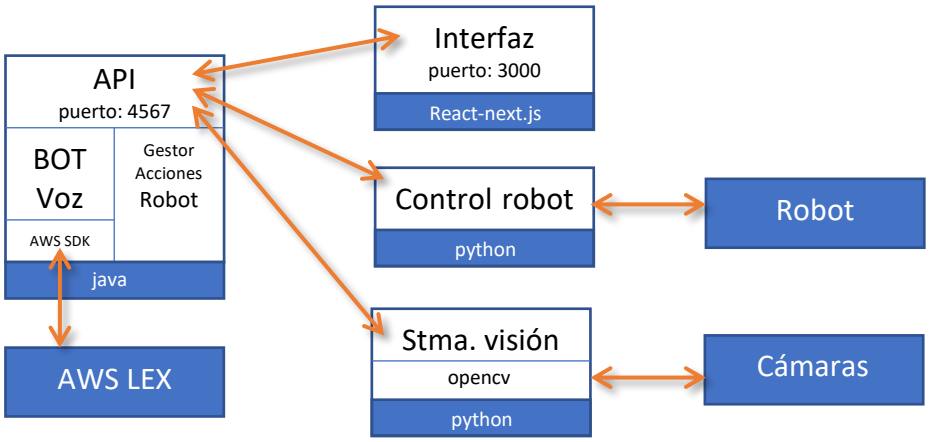
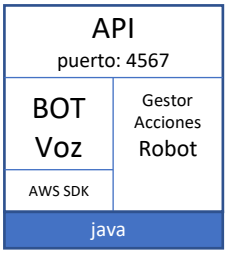


Figura 8: esquema de componentes del sistema completo
Fuente: elaboración propia

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

El módulo principal del sistema, realizado con Java consta de los siguientes elementos principales:



API: se publica un API en el puerto 4567 donde pueden acceder:

- el interfaz para activar el control de voz, mostrar los comandos y el estado del sistema
- el sistema de control del robot, para acceder a los comandos que tiene que ejecutar el robot e indicar el estado del robot
- el sistema de visión, el cual tiene que publicar un json con los datos de posicionamiento de los objetos tanto en el almacena como en la zona del usuario.

BOT de VOZ: Se encarga de comunicarse, a través del AWS SDK con el AWS Lex, conectarse con el bot creado en AWS, transmitirle el audio, recibir las respuestas de AWS Lex, interpretarlas y crear los comandos.

Gestor de acciones del ROBOT: A través de los comandos creados por el BOT de VOZ, se crean las órdenes a realizar por el robot, accediendo a los datos recibidos por el sistema de visión para detectar la posición de los distintos elementos que componen los comandos. Estos comandos se van encolando y se van ofreciendo al Control del Robot, según se vayan pidiendo. Se recibe el estado del robot y de las órdenes

Pantallazos de la programación del módulo principal:

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

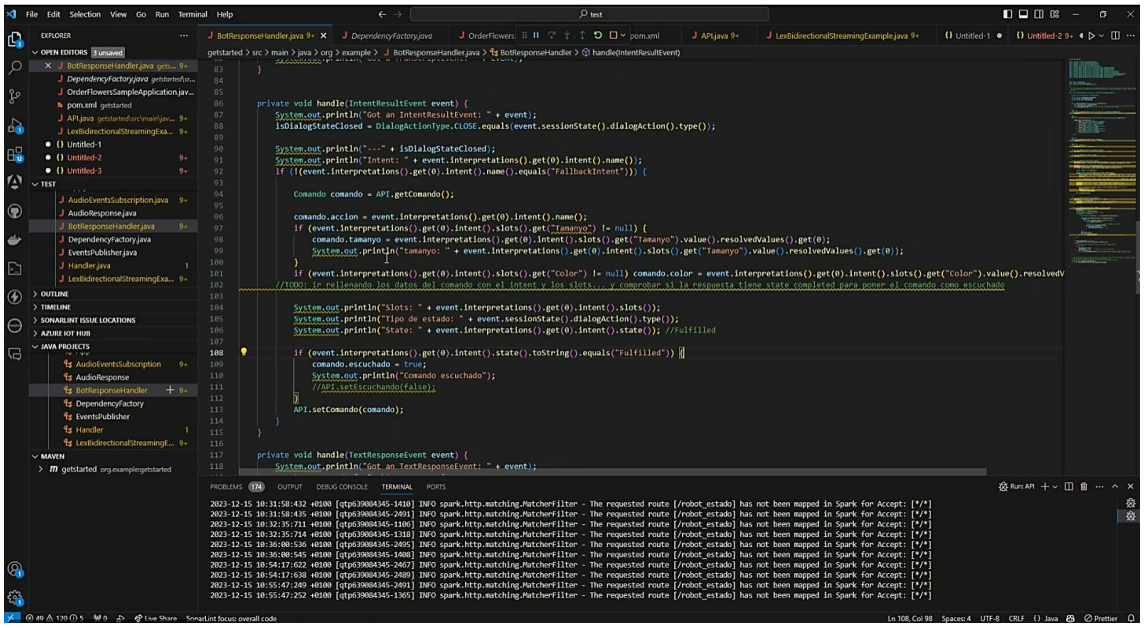


Figura 9: Ejemplo de programación generada
Fuente: elaboración propia.

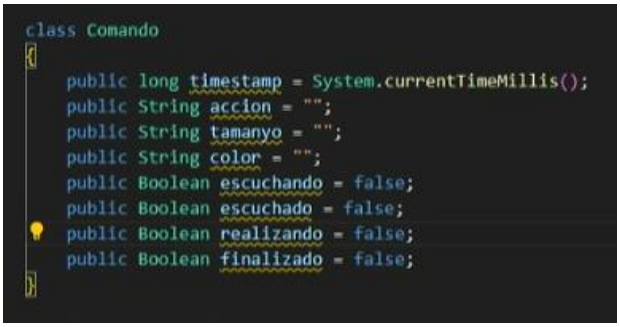
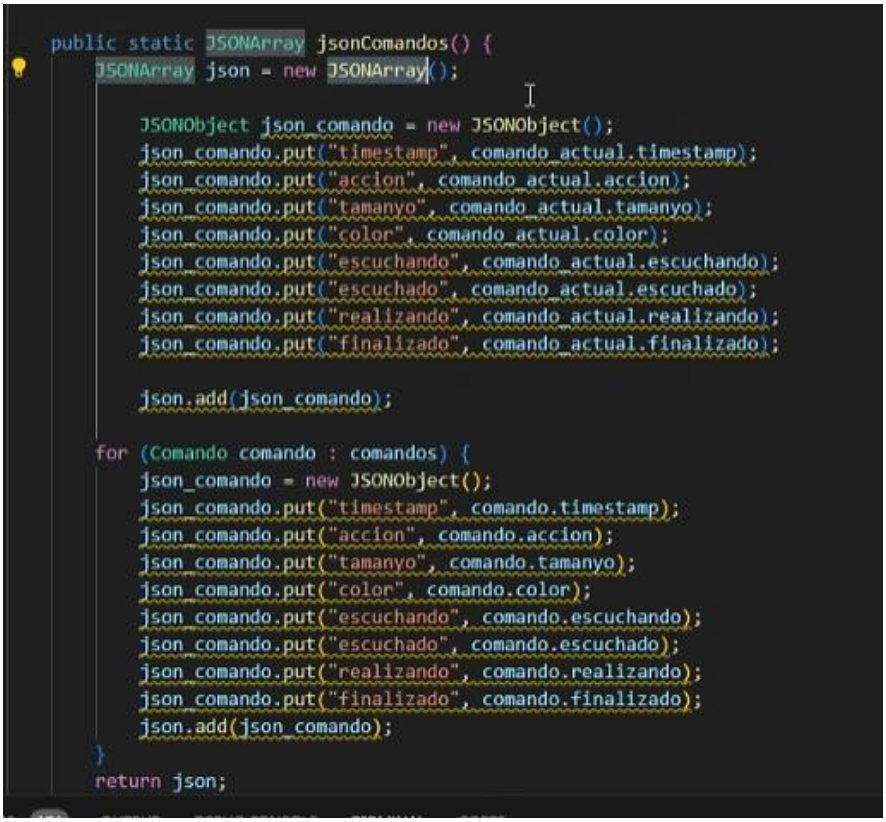


Figura 10: Detalle de programación generada (i)
Fuente: elaboración propia.

AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea



```
public static JSONArray jsonComandos() {
    JSONArray json = new JSONArray();

        JSONObject json_comando = new JSONObject();
        json_comando.put("timestamp", comando_actual.timestamp);
        json_comando.put("accion", comando_actual.accion);
        json_comando.put("tamanyo", comando_actual.tamanyo);
        json_comando.put("color", comando_actual.color);
        json_comando.put("escuchando", comando_actual.escuchando);
        json_comando.put("escuchado", comando_actual.escuchado);
        json_comando.put("realizando", comando_actual.realizando);
        json_comando.put("finalizado", comando_actual.finalizado);

        json.add(json_comando);

    for (Comando comando : comandos) {
        json_comando = new JSONObject();
        json_comando.put("timestamp", comando.timestamp);
        json_comando.put("accion", comando.accion);
        json_comando.put("tamanyo", comando.tamanyo);
        json_comando.put("color", comando.color);
        json_comando.put("escuchando", comando.escuchando);
        json_comando.put("escuchado", comando.escuchado);
        json_comando.put("realizando", comando.realizando);
        json_comando.put("finalizado", comando.finalizado);
        json.add(json_comando);
    }
    return json;
```

Figura 11: Detalle de programación generada (ii)
Fuente: elaboración propia.

AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea

2.3.7 Desarrollo interfaz de usuario

Se ha desarrollado una interfaz de usuario, utilizando para ello react-next.js, un framework basado en react, orientado para la realización de aplicaciones de front-end.

Desde este desarrollo se puede acceder al API del módulo principal, ver los comandos que se han almacenado, el estado del comando actual que se está reconociendo por el sistema, el estado en el que se encuentra el resto de los comandos, controlar la puesta en marcha del sistema de voz y ver el estado del robot y el sistema de control.

Desde este interfaz, se puede gestionar la parada o el inicio del sistema de control de voz, y se visualiza la información de lo que está pasando con el sistema de voz, el sistema de visión y el robot.

Aquí el usuario debe de conocer:

- Si se han interpretado correctamente su orden verbal.
- Si se ha transmitido
- Si el robot lo está ejecutando
- Cuando se ha terminado de ejecutar una acción

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

Figura 12: Diseño de interfaz de usuario
Fuente: elaboración propia.

2.4 Tarea 4.4 Desarrollo de pruebas

El objetivo de las pruebas se puede resumir en:
- Verificar la eficacia del sistema para reconocer e interpretar la INTENTION formulada por un usuario (entregar o devolver una pieza), el COLOR (rojo, amarillo o azul) y el TAMAÑO de la pieza (grade o pequeño).

Para ello, se han llevado a cabo pruebas preliminares de funcionamiento del sistema desarrollado emitiendo diferentes instrucciones verbales que siguen la siguiente estructura, expresada en lenguaje natural:

<INTENTION> + <COLOR> + <TAMAÑO>

Algunos ejemplos de frases de testeo han sido:

“Por favor, me puedes entregar una pieza amarilla pequeña”
“Dame una pieza azul grande”
“Devuelve un bloque rojo grande”
“Me gustaría devolver una pieza de color amarilla pequeña”

Las pruebas han sido llevadas a cabo por diferentes usuarios (técnicos de AIDIMME participantes en el proyecto).

3 Resultados obtenidos

Durante la realización de las pruebas, se ha reentrenado al sistema con nuevas expresiones y palabras, tras detectar en las primeras iteraciones errores en la interpretación de las órdenes verbales emitidas. Este proceso de pruebas y entrenamiento del sistema ha tenido varias iteraciones hasta comprobar que era fiable dentro del dominio entrenado.

Se ha incidido especialmente en aportar al sistema sinónimos o expresiones que complementan las palabras clave. Por ejemplo:

- Para la intention “dar”: entregar, acercar, etc.
- Para “color”: la de color amarillo, la pieza roja, etc.

Se detectó inicialmente que el sistema no reconocía verbos con pronombres reflexivos del estilo: acércame, tráeme, llévate, etc. Por tanto, se volvió a entrenar al sistema con este tipo de expresiones de forma explícita.

- Para la intention “dar”: entrégame, acércame, llévate, etc.

A la hora de definir los indicadores que permitieran validar este sistema dentro del demostrador del paquete de trabajo PT6, se contó con la participación de las empresas colaboradoras en el proyecto.

En diferentes reuniones se les encuestó acerca de que indicadores consideraban importantes para medir la calidad del sistema de interpretación del lenguaje natural desarrollado, obteniendo el siguiente el resultado.

Cuestión T4.4a	HURTADO R.	CFZ	DYMSA	Promedio
Tasa de error de interpretación de la acción completa (“tráeme la pieza roja grande”)	4	5	5	**4,7**
Tasa de error de interpretación de un componente de la acción (color pieza, tamaño pieza, …)	5	4	5	**4,7**
Tiempo de procesamiento de la orden verbal	3	3	5	**3,7**

Figura 13: Tabla de cuestiones T4.4a
Fuente: elaboración propia.

AIDIMME INSTITUTO TECNOLÓGICO · GENERALITAT VALENCIANA · IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN · Financiado por la Unión Europea

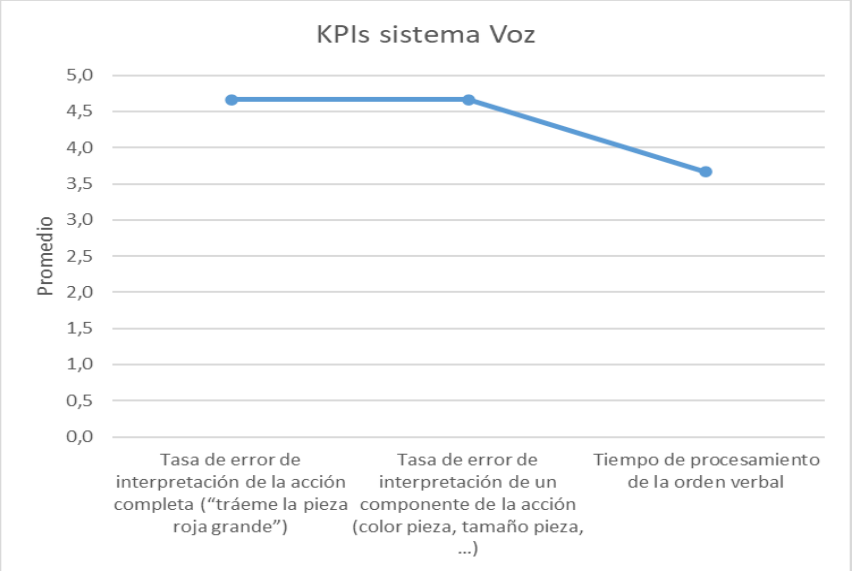


Figura 14: Grafica promedio T4.4a
Fuente: elaboración propia.

Las valoraciones se realizaron bajo una escala Likert entre 1 y 5, donde 1 implicaba la mínima importancia y 5 la máxima.

Se consideró que la tasa de interpretación de la acción completa y de un componente de la acción (color o tamaño) era más relevante que el tiempo de interpretación de la acción por parte del sistema.

Además, se encuestó acerca de la importancia de diferentes parámetros que podían influir en el buen funcionamiento de este sistema, obteniendo los siguientes resultados siguiendo el mismo sistema de escala Likert.

Cuestión T4.4b	HURTADO R.	CFZ	DYMSA	Promedio
Nivel de ruido industrial (no humano) en el entorno	5	5	2	**4,0**
Nivel de ruido de otras voces humanas en el entorno	4	5	4	**4,3**
Volumen del tono de voz emitido	1	2	2	**1,7**
Velocidad de emisión de órdenes verbales	1	3	3	**2,3**
Acentos regionales del emisor de voz	2	3	4	**3,0**

Figura 15: Tabla de cuestiones T4.4b
Fuente: elaboración propia.

AIDIMME INSTITUTO TECNOLÓGICO GENERALITAT VALENCIANA IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN Financiado por la Unión Europea

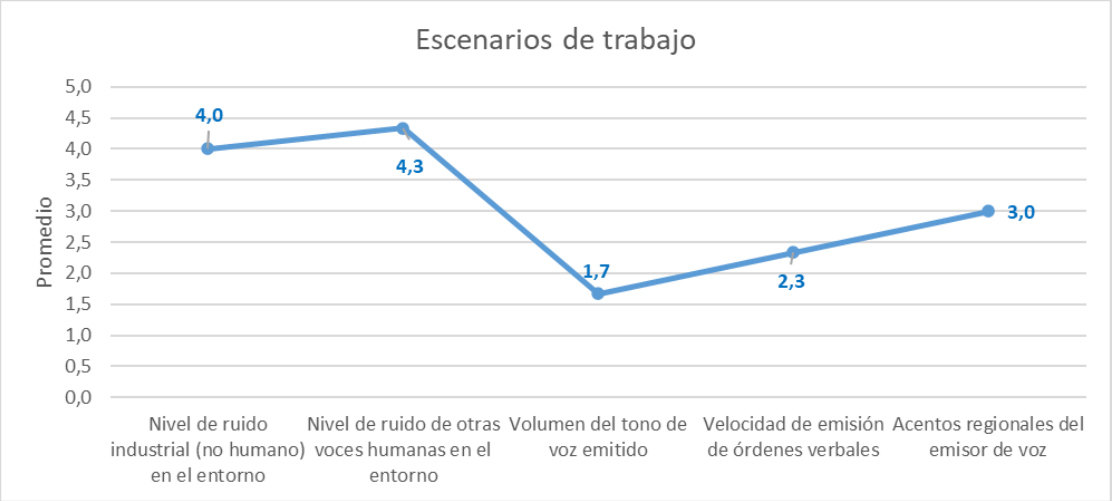


Figura 16: Gráfica de resultados T4.4b
Fuente: elaboración propia.

El resultado fue que el nivel de ruido, tanto industrial como de otras voces humanas eran los dos parámetros que mayor influencia podían presentar.

Estos resultados han sido utilizados como input dentro del trabajo de definición del piloto demostrador y del diseño de las pruebas de validación en el PT6.

4 Conclusiones

Como conclusiones al desarrollo de las actividades ejecutadas en el paquete de trabajo PT4 asociado al presente entregable podemos destacar las siguientes:

- Se ha generado una ontología para la interacción entre humano-robot, que se ajusta al caso de uso que se va a desarrollar en las pruebas de validación que se llevan a cabo en el PT6.

- Esta ontología especifica de forma formal y explícita los conceptos relevantes en el dominio de las relaciones humano-robot, y supone un punto de partida que puede ser enriquecido con nuevos casos de uso diferentes al del presente proyecto.

- Se ha generado un sistema que interpreta el lenguaje natural emitido por voz de un usuario que solicita acciones a un robot.

- Este sistema se ha desarrollado sobre la tecnología de AWS Lex para un caso de uso donde el robot debe manipular piezas de dos tamaños y tres colores.

- El sistema se ha testado con diferentes usuarios. Tras varias iteraciones de reentrenamiento, la fiabilidad se ha demostrado como elevada. Esta fiabilidad es evaluada en el paquete de trabajo PT6 donde se monta el demostrador del proyecto y se testará con diferentes usuarios y condiciones ambientales de ruido.

AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea

5 Anexos y bibliografía

Neches, Robert & Fikes, Richard & Finin, Tim & Gruber, Thomas & Patil, Ramangouda & Senator, Ted & Swartout, William. (1991). Enabling Technology for Knowledge Sharing.. AI Magazine. 12. 36-56.

Guarino, Nicola. (1998). Formal Ontologies and Information Systems.

Vilches-Blázquez, Luis M. & García-Silva, Andrés & Villazon Terrazas, Boris. (2009). Construcción de ontologías a partir de tesauros.

AIDIMME INSTITUTO TECNOLÓGICO
GENERALITAT VALENCIANA
IVACE+i INSTITUTO VALENCIANO DE COMPETITIVIDAD E INNOVACIÓN
Financiado por la Unión Europea

AIDIMME
INSTITUTO TECNOLÓGICO

Domicilio fiscal —
C/ Benjamín Franklin 13. (Parque Tecnológico)
46980 Paterna. Valencia (España)
Tlf. 961 366 070 | Fax 961 366 185

Domicilio social —
Leonardo Da Vinci, 38 (Parque Tecnológico)
46980 Paterna. Valencia (España)
Tlf. 961 318 559 - Fax 960 915 446

aidimme@aidimme.es
www.aidimme.es