

MEGARA

Servinform ejecuta desde enero 2023 hasta junio 2024 el PROYECTO I+D MEGARA. Plataforma de modelos lingüísticos basada en grandes transformadores preentrenados para la comprensión del lenguaje. (Linguistic ModEling platform based on larGe pretrAined tRansformers for lAnguage understanding)

- Términos Clave: Automatización del Lenguaje, Procesamiento del Lenguaje Natural, Inteligencia Artificial, Aprendizaje Automático, Aprendizaje Profundo, Modelos Largos de Lenguaje, Transformadores.
- Topics: Language Automation, Natural Language Processing (NLP), Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Large Language Models (LLM), Transformers.

“Si tuviera ocho horas para talar un árbol, dedicaría seis a afilar mi hacha.”

Con el proyecto MEGARA se aspira a obtener una plataforma tecnológica de automatización del lenguaje, que integrando tecnologías Natural Language Processing, Deep Learning, Large Language Models, Transformers, Transfer Learning, Few-shot learning, permita ofrecer a todos los sectores de actividad empresarial servicios tecnológicos avanzados de automatización en el tratamiento del lenguaje, integrables en los procesos de gestión empresarial, y que potencien la automatización de estos.

Por ‘automatización del lenguaje’, entendemos la práctica de servicios informáticos o computacionales que permitan interactuar lingüísticamente de forma automatizada con personas, en un entorno de procesos de gestión empresarial.

Por lo tanto, dentro del ámbito de la automatización del lenguaje, consideramos como casos de uso o procesos de interés los siguientes a título ilustrativo, no exhaustivo:

I. Verificación de texto: corrector ortográfico, corrector gramatical, etc.

II. Análisis automatizado de sentimiento en comentarios de redes sociales.

III. Clasificación email, no sólo anti-spam sino por ejemplo en base al departamento que lo ha de gestionar.

- IV. Filtrado de emails salientes desde las cuentas de email de una organización.
- V. Tipificación y respuesta automatizada a incidencias, peticiones y solicitudes a través de canales digitales.
- VI. Clipping automatizado de medios on line de prensa, blogs, redes sociales, etc.
- VII. Análisis en tiempo real de feeds en redes sociales y flujos de texto para emisión de alertas ante la detección de determinados patrones.
- VIII. Extracción automatizada de entidades de datos de textos complejos y extensos de un determinado dominio específico: jurídico, técnico, etc.
- IX. Motores de búsquedas personalizados para un determinado contexto o dominio de conocimiento.
- X. Detectores de plagio mediante comparaciones aproximativas, con detección de contenido parafraseado.
- XI. Generadores de corpus lingüísticos partiendo de fuentes de datos diversas.
- XII. Chatbot conversacional automatizado.
- XIII. Generación de imágenes partiendo de un texto descriptivo.
- XIV. Etc.

Los grandes retos de la automatización del lenguaje pasan por:

A. Las dificultades para integrar conocimiento experto de la ingeniería informática y conocimiento experto de la lingüística.

B. Las dificultades y la complejidad de obtener y preparar adecuadamente datos de entrenamiento y ajuste para las soluciones de automatización del lenguaje.

C. El coste del ajuste y entrenamiento de los modelos de lenguaje avanzados.

D. La replicabilidad de la respuesta en modelos de lenguaje públicos compartidos.

E. La evaluación de la fiabilidad en cada caso de un proceso concreto de tratamiento automatizado del lenguaje.

F. La complejidad derivada del multilingüismo en cualquier tratamiento automatizado del lenguaje que comprenda un tratamiento simultáneo de texto en diversos idiomas.

G. La calidad del dato y la creciente degradación del correcto uso del lenguaje por parte de las personas.

H. La dependencia de la conectividad a Internet para el funcionamiento de los más potentes modelos de tratamiento de lenguaje.

I. Las dificultades específicas en el tratamiento de textos muy largos.

A todos y a cada uno de estos retos se aspira a dar algún nivel de respuesta desde el proyecto MEGARA.

El objetivo principal del proyecto MEGARA es la investigación y el desarrollo para la obtención de una solución de tecnología que permita avanzar lo máximo posible en los siguientes puntos:

A. Identificar y abstraer las tareas más habituales y de interés para la automatización computacional persona-máquina de la interlocución lingüística, en el ámbito de la gestión de los procesos empresariales.

B. Potenciar el tratamiento avanzado automatizado en todo tipo de interlocución lingüística persona-máquina en el ámbito de la gestión de los procesos empresariales, aplicando soluciones basadas en módulos y componentes AI y NLP.

C. Definir e implementar módulos y componentes AI/NLP autocontenidos y autosuficientes en la realización de una acción específica dentro del propio proceso. Esto aportará un grado general de eficiencia y eficacia.

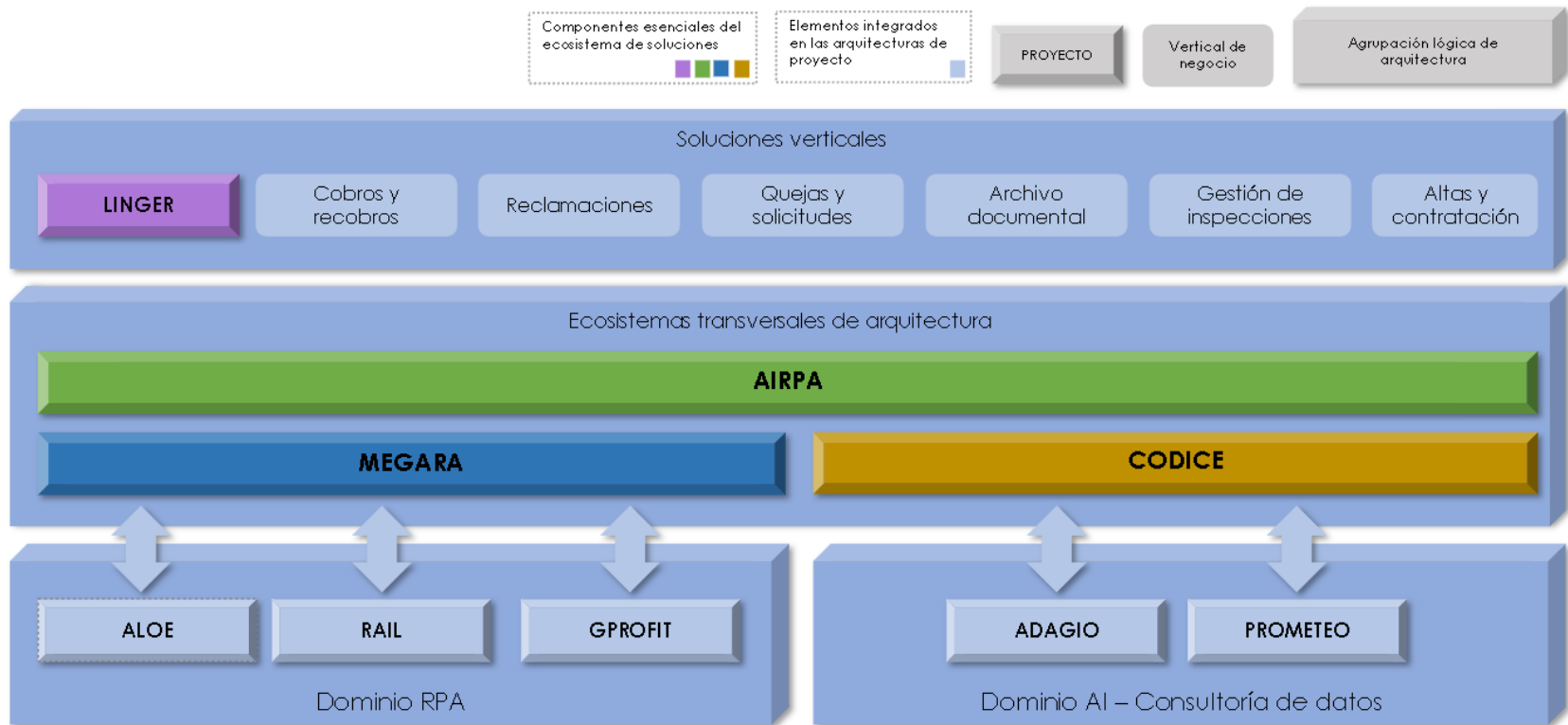
D. Implementar nuevas técnicas de interlocución lingüística potenciadas con tecnologías innovadoras, con el objeto de reducir la participación humana directa en la ejecución de dichas tareas, y así también reducir las tareas residuales que no aportan valor y los retrabajos que el propio proceso actualmente asume.

E. Ofrecer un framework lo más independiente posible de productos software concretos, soluciones particulares, o tecnologías concretas, que ayude desde una perspectiva global y general a superar las dificultades en la automatización de las tareas más habituales de interlocución lingüística persona-máquina.

F. Los resultados deben pasar por intervenir en el proceso de forma transparente, mínimamente intrusiva, y eficiente, creando puntos de unión y coordinación entre los nuevos flujos de trabajo propuestos, y el proceso actual en la situación de partida.

MEGARA

MEGARA en la estrategia I+D+i del Área de Consultoría e Innovación de Servinform



MEGARA

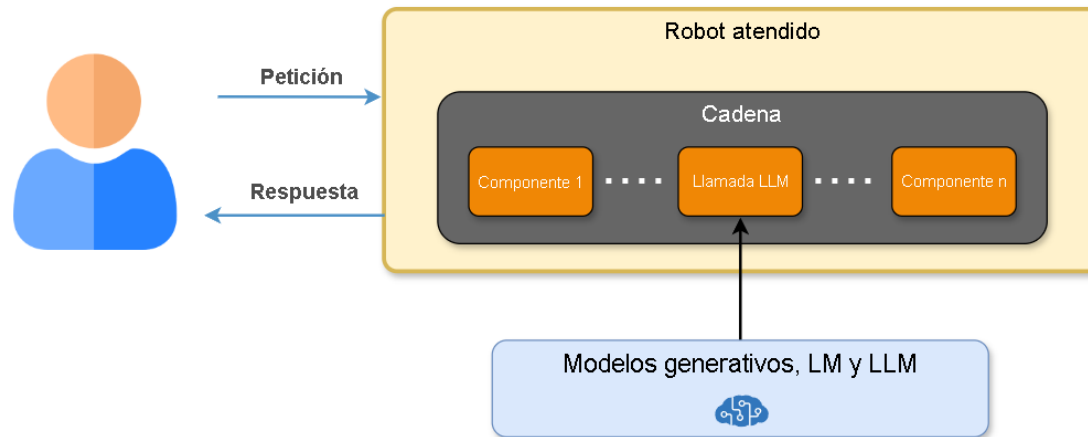
Ciclo de vida de la información



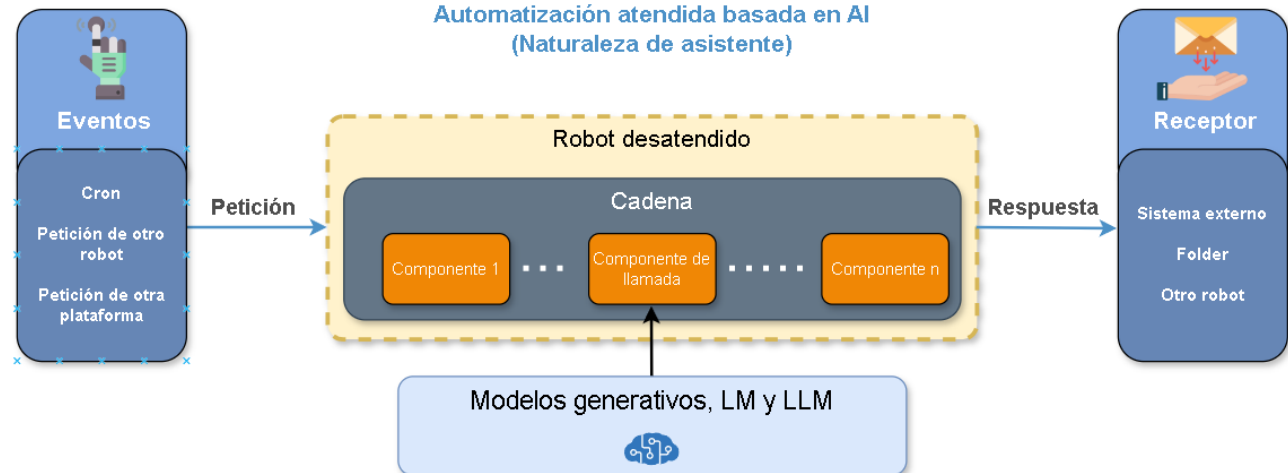
MEGARA

Automatización atendida y desatendida MEGARA

Automatización atendida basada en AI (Naturaleza de asistente)

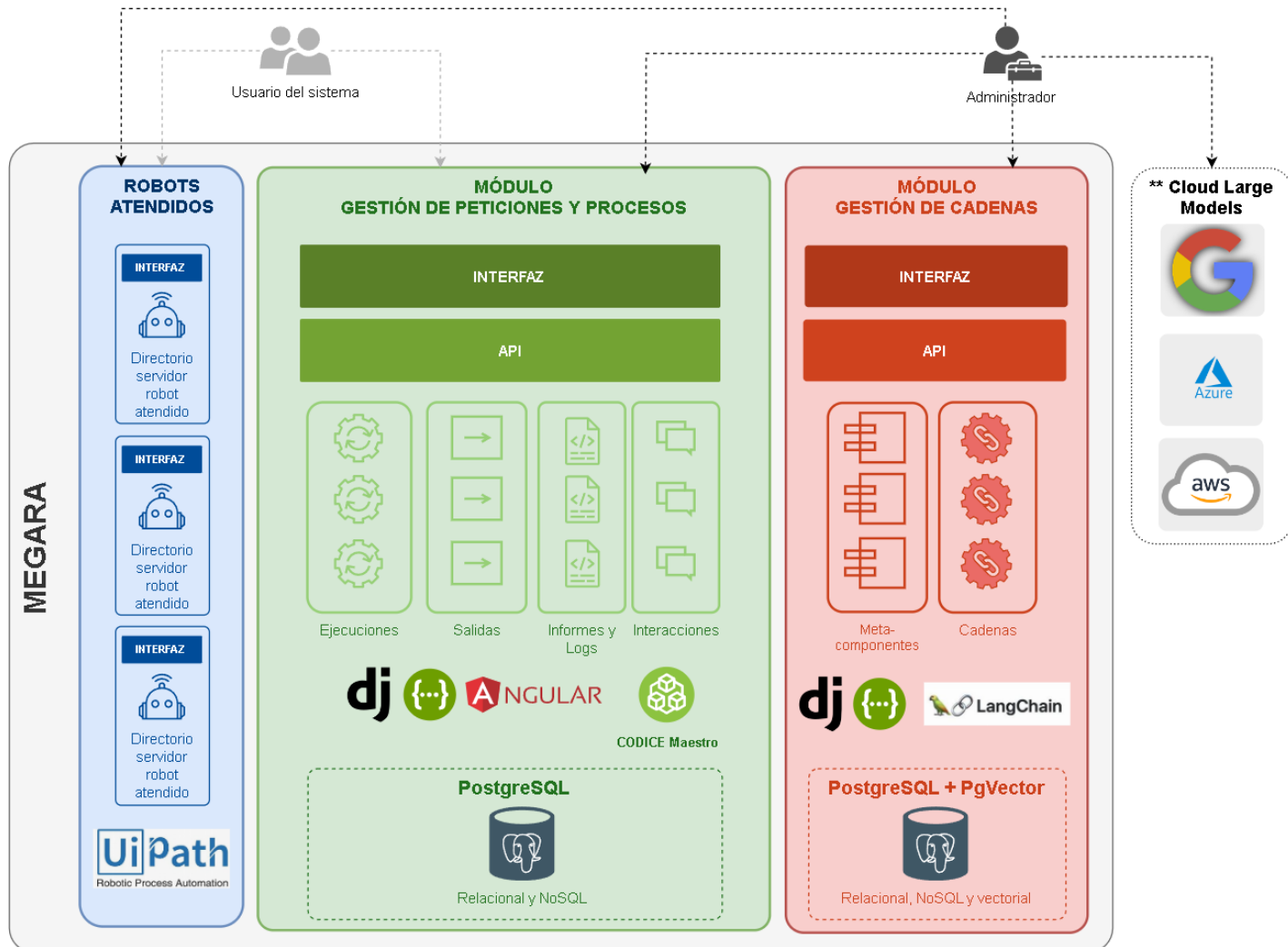


Automatización atendida basada en AI (Naturaleza de asistente)



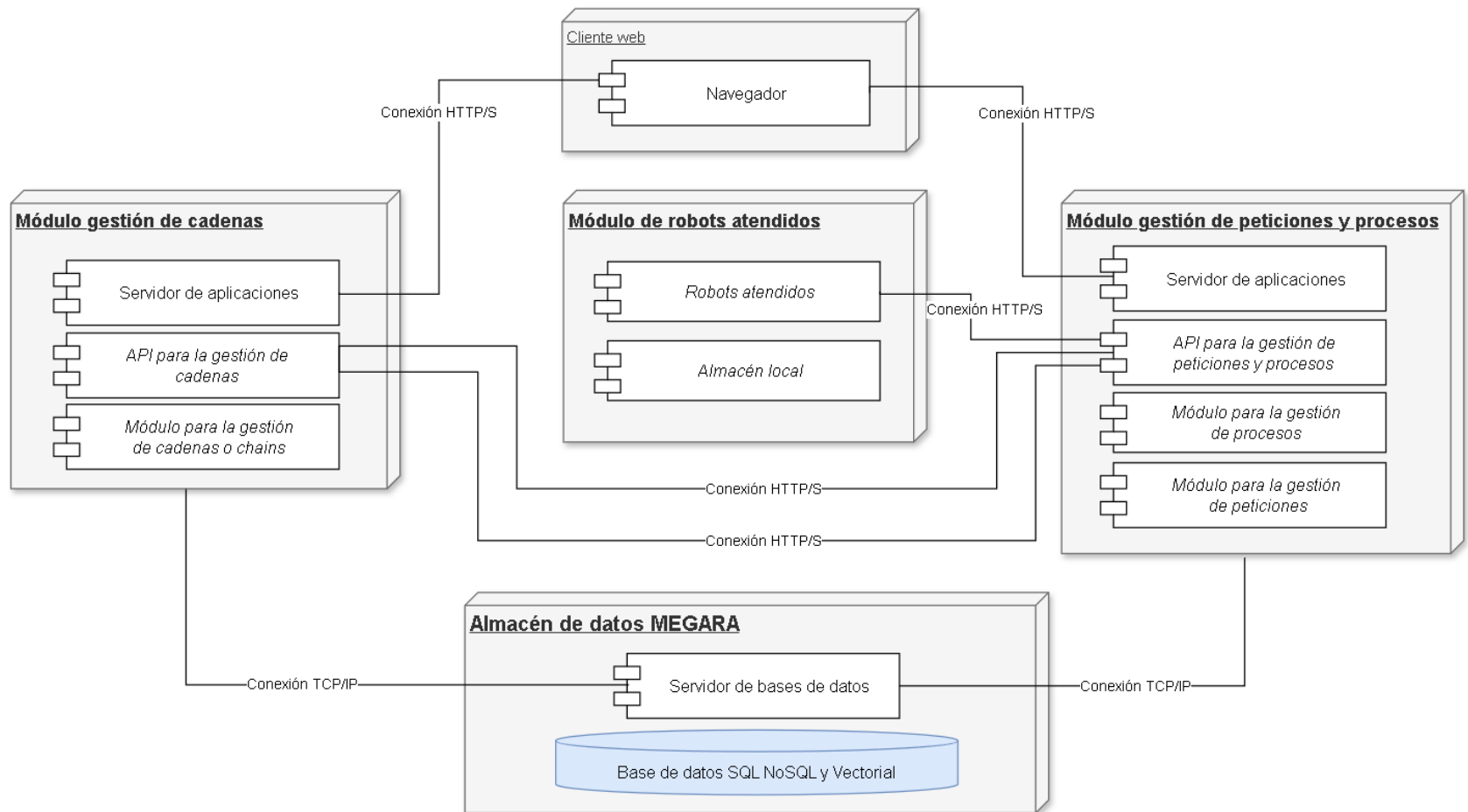
MEGARA

Arquitectura MEGARA



MEGARA

Diagrama de componentes de la selección





CDTI - CENTRO PARA EL DESARROLLO TECNOLÓGICO INDUSTRIAL
SECRETARÍA DE ESTADO DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD
GOBIERNO DE ESPAÑA

Más información en Internet en:

<https://www.servinform.es/soluciones-tecnologicas/>