

Sistema automático de consultas mediante IA implementado en un servidor de bajas prestaciones

MARTIN, Raúl Lisandro¹; CREPALDO, Daniel²; PACHER³, Federico
VARELA⁴, Carlos; GHORGHOR⁵, Javier; BAILON⁶, Eduardo.

Laboratorio de Microelectrónica – FCEIA - UNR

¹rlmartin@fceia.unr.edu.ar

³fpacher@fceia.unr.edu.ar

⁵javghor@fceia.unr.edu.ar

²crepaldo@fceia.unr.edu.ar

⁴charlyv@fceia.unr.edu.ar

⁶ebailon@fceia.unr.edu.ar

Palabras claves: Inteligencia Artificial. Educación. Consultas. Accesible..

INTRODUCCIÓN

Durante la última pandemia se puso en evidencia la importancia de las herramientas virtuales para llevar la enseñanza a lugares que en esa situación eran inaccesibles. Estas herramientas pueden también ser útiles en el contexto presente permitiendo a los estudiantes disponer de múltiples opciones a la hora de comunicarse con los docentes, obtener información, realizar consultas, etc.

En este marco, surgió la inquietud de explorar cómo las nuevas tecnologías de procesamiento de lenguaje natural, específicamente los modelos de lenguaje de gran escala (LLMs [1]), podrían integrarse a contextos educativos reales con recursos limitados. La propuesta se centró en resolver una necesidad concreta: permitir que los estudiantes puedan realizar consultas académicas específicas, basadas exclusivamente en la bibliografía oficial, sin depender de horarios preestablecidos ni de la disponibilidad del cuerpo docente. A su vez, se buscó intencionalmente evitar el uso de servicios en la nube o hardware especializado, lo que permitiría a instituciones educativas con presupuestos acotados implementar soluciones de IA funcionales y accesibles.

METODOLOGÍA / MATERIALES Y MÉTODOS

Se aplicó para el desarrollo de este proyecto la metodología de investigación-acción. La misma implica métodos que incluyen el objetivo de investigar, y el de actuar/modificar algo de lo investigado. A partir de la información recopilada surge un nuevo plan de acción, configurando una estrategia recursiva.

DESARROLLO

Todo el conjunto se instaló en una PC de escritorio con las siguientes características:

- Procesador Intel I3 3a generación (Intel(R) Core(TM) i3-4170 CPU @ 3.70GHz).
- Disco rígido de estado sólido (SSD) de 400 Gb de capacidad.
- Memoria RAM DDR3 de 16 Gb.
- Sistema operativo Linux distribución Ubuntu 24.04.2 LTS.

Una vez instalado el sistema en la PC fue sometido a una batería de pruebas para verificar su funcionamiento y efectuar las correcciones del código necesarias.

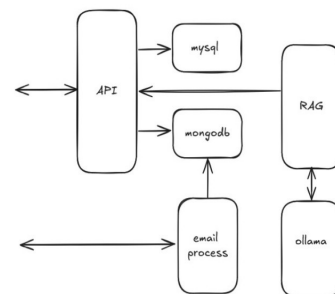


Figura 1. Diagrama en bloques

RESULTADOS y DISCUSIÓN

El sistema se encuentra funcionando en forma permanente desde su implementación, con documentación de una materia de tercer año de la carrera de Ingeniería Electrónica, siendo sometido a consultas periódicamente. En la figura 2 puede verse el formato de respuesta elegido, el cual es incorporado automáticamente al correo de salida que contiene la respuesta generada por el sistema.

Las respuestas del sistema resultan en su gran mayoría satisfactorias, si bien en algunos casos aparecen algunos inconvenientes con el lenguaje, como reemplazar la palabra "juntura" por "junción".

El tiempo inicial de de las consultas fue de 40 minutos. Posteriormente se realizaron mejoras en cuanto a la recuperación de documentos, modificación de cantidad de tokens (de recuperación, de generación y overlap), mediante lo cual se ha llegado a reducir el tiempo de respuesta a la mitad.

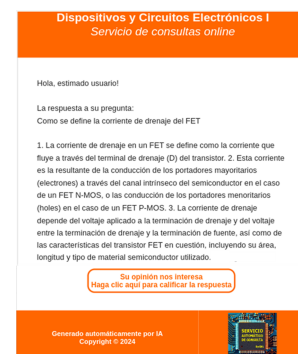


Figura 2. Formato de respuesta

CONCLUSIONES

Se desarrolló un sistema que puede dar a los estudiantes un soporte académico que complementa a las instancias de consultas tradicionales, permitiéndoles realizar consultas a demanda asincrónica, en cualquier momento de manera automática, específica y accesible desde su celular, sin necesidad de un hardware o software especializado. Este sistema está implementado en una PC de bajos requerimientos lo que permite su utilización sin efectuar grandes inversiones, por ejemplo en instituciones educativas públicas.

AGRADECIMIENTOS y FINANCIAMIENTO (OPCIONAL)

Este trabajo se generó a partir de la propuesta "Aplicación de un modelo de lenguaje (LLM) a un uso educativo implementado en servidores de pequeño porte", la cual fue presentada en 2024 a la convocatoria XperienciaUNR de la Universidad Nacional de Rosario. Dicha propuesta fue seleccionada por lo que la UNR brindó soporte y subvencionó la totalidad de su desarrollo