

Este Trabajo de Fin de Grado con nombre “LLMSearch: Buscador multimedia basado en lenguaje natural”, consiste en desarrollar un sistema que permita la búsqueda de archivos multimedia mediante consultas formuladas en lenguaje natural como si se le preguntase a una persona. La motivación del proyecto surge de la dificultad para localizar archivos específicos dentro de grandes volúmenes de datos, en especial cuando el usuario solo recuerda detalles parciales del contenido buscado.

El problema principal es que los sistemas de búsqueda tradicionales dependen exclusivamente de nombres de archivo exactos o metadatos específicos, lo cual resulta insuficiente en la gran mayoría de casos cuando se quiere buscar un archivo específico.

Para abordar este problema, se ha desarrollado una solución basada en Inteligencia Artificial multimodal, capaz de procesar simultáneamente textos, imágenes y otro tipo de formatos más avanzados como audio o vídeo.

La arquitectura implementada sigue el paradigma Retrieval-Augmented Generation (RAG) el cual organiza el proceso de búsqueda en 2 partes. Primero, se recuperan los archivos más relevantes mediante embeddings vectoriales a partir de la consulta del usuario sobre la base de datos. Posteriormente, se genera una respuesta adaptada utilizando modelos de lenguaje natural y, en este caso, un prompt específico que permite filtrar y modificar, en caso de ser necesario, dicha respuesta de la base de datos.

El proyecto se ha estructurado siguiendo una adaptación simplificada de la metodología Scrum, dividiendo el trabajo en sprints iterativos de aproximadamente 2 semanas.

La implementación técnica combina diversas herramientas modernas. Vue.js proporciona la interfaz de usuario mientras que Flask en Python gestiona la API REST del backend. Prefect se encarga de la orquestación de tareas y LMStudio facilita la ejecución local de modelos de lenguaje cuantizados. Esta arquitectura modular garantiza escalabilidad y facilidad de mantenimiento.

Una característica importante del sistema es su capacidad de poder ejecutarse de manera completamente local, de esta manera el usuario puede utilizar este sistema de forma privada y segura. Además, si el usuario no dispusiera de un dispositivo con suficiente rendimiento para utilizar los modelos en local estos podrían usarse desde la nube. Pero la idea de que el sistema sea modular es que el usuario pueda utilizar un modelo pequeño y optimizado para su dispositivo.

Las evaluaciones realizadas demuestran que el sistema logra identificar archivos relevantes mediante consultas en lenguaje natural, ofreciendo resultados considerablemente más precisos que los métodos de búsqueda tradicionales. El rendimiento se mantiene estable incluso en dispositivos de uso doméstico, pero esto depende de los modelos utilizados.

Los resultados obtenidos confirman que aplicar Inteligencia Artificial sobre este problema es viable. El proyecto no solo ofrece una solución funcional a un problema cotidiano, sino que también establece fundamentos para futuras investigaciones en búsqueda multimedia inteligente.

En conclusión, este proyecto demuestra como la Inteligencia Artificial puede ayudar de manera considerable en la búsqueda de contenido multimedia, acercando más la tecnología al ámbito doméstico y cotidiano de los usuarios.