

Proyecto 3:

Introducción:

Este proyecto utiliza inteligencia artificial para analizar archivos en formato Markdown (.md) provenientes de la Junta de Andalucía en el ámbito educativo. Su objetivo principal es entender el significado detrás del texto y ofrecer respuestas pertinentes a las preguntas formuladas sobre estos documentos. El procesamiento del lenguaje natural (NLP) es esencial para lograr este análisis semántico de manera efectiva.

Uso del Procesamiento del Lenguaje Natural (NLP):

Extracción de Información:

Identificamos los temas principales y la información relevante en los documentos .md.

Recopilamos datos específicos como fechas, nombres, eventos, etc.

Generación de Respuestas:

Usamos modelos de lenguaje para crear respuestas coherentes y pertinentes a las preguntas.

Análisis Semántico:

Entendemos el significado y contexto de las palabras y frases para mejorar la calidad de las respuestas.

Transformación de Formato:

Convertimos archivos PDF a Markdown para facilitar el procesamiento y análisis de texto.

Modelo de Lenguaje (Mixtral-8x7B-Instruct-v0.1):

Usamos el modelo Mixtral-8x7B-Instruct-v0.1 para comprender instrucciones y contextos complejos en tareas de procesamiento del lenguaje natural.

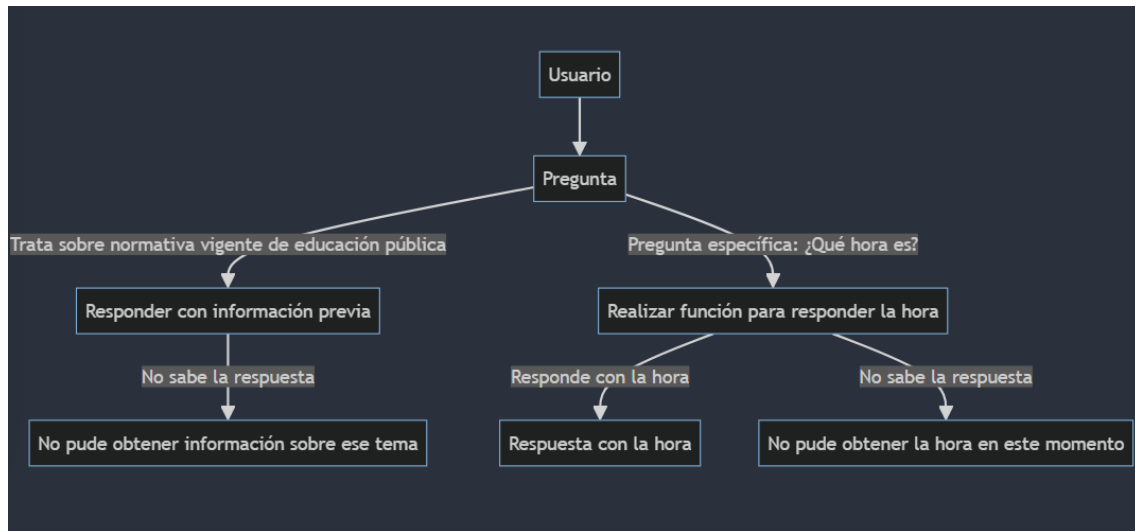
Embedding (BAAI/bge-large-en):

Utilizamos el embedding proporcionado por BAAI/bge-large-en para representar palabras como vectores. Esto ayuda al modelo a entender mejor las relaciones entre las palabras en el texto.

Despliegue en Gradio:

La interfaz de usuario y la implementación del modelo se hacen utilizando Gradio, una biblioteca que simplifica la creación de interfaces interactivas para modelos de aprendizaje automático.

Diagrama Conversacional (Mermaid):



Conclusión:

En resumen, la implementación del procesamiento del lenguaje natural en este proyecto ofrece una comprensión más profunda y precisa de los documentos educativos recibidos. La utilización de modelos avanzados, embeddings específicos y herramientas de despliegue contribuye a una solución eficiente y efectiva para responder a consultas sobre estos documentos. La aplicación de NLP se muestra como esencial para extraer información valiosa y proporcionar respuestas contextualmente relevantes.