

Aplicaciones LLM

Componentes técnicos de RAG

Recuerda: RAG supera context window

- Preparativos:
 - Divide tus datos en pequeños segmentos.
 - Convierte los pequeños segmentos en números (“embeddings”).
 - Carga los embeddings en una base de datos vectorial.
- Ahora cuando preguntas (“query”) al LLM:
 - El LLM va a la base de datos vectorial y busca (“indexing”) solo (“semantic similarity”) los datos que responden a tu pregunta.

Recuerda: Importancia de la técnica RAG

- RAG es esencial para crear Aplicaciones LLM.
- Por eso nos detendremos en dominar esta técnica.

Componentes técnicos de RAG

- A lo largo del programa los veremos en detalle:
 - Embeddings
 - Vector Databases
 - Semantic Similarity
 - Query
 - Indexing
 - Orchestration