

Asesor Legal para Consulta de Historia de Demandas

Prueba Técnica - Natalia Baez De La Luz

El Problema del Negocio

Situación Actual

329 casos almacenados en Excel con búsqueda manual fila por fila

Impacto

Consume tiempo valioso de los abogados y dificulta respuestas rápidas a clientes

Necesidad

Automatizar el primer nivel de consulta para enfoque en casos complejos

Alcance de la Prueba de Concepto

Foco Principal

Demandas relacionadas con **redes sociales**

Objetivo


Respuestas automáticas en lenguaje claro y comprensible

Preguntas Ejemplo

- ¿Cuáles son las sentencias de 3 demandas?
- ¿Cuál fue la sentencia del caso sobre acoso escolar?
- ¿Existen casos que hablen sobre el PIAR, y cuáles fueron sus sentencias?

Evaluación de Alternativas Técnicas

Enfoque	Ventajas	Desventajas	Complejidad
Búsqueda por palabras clave	Rápido, simple	No entiende sinónimos	Muy baja
TF-IDF / BM25 TF (Term Frequency) IDF (Inverse Document Frequency)	Mejor que búsqueda básica	No captura significado semántico. sigue siendo una búsqueda lexical. No entiende el sentido del texto.	Baja
RAG (Elegido)	Búsqueda semántica + respuestas naturales	Mayor complejidad técnica	Media/alta

 **Decisión:** RAG permite entender que "bullying" = "acoso escolar" y genera respuestas en lenguaje natural

Solución: Arquitectura RAG

RAG = Retrieval-Augmented Generation

Recuperación + Generación

1

Fase de Preparación

Cargar 329 casos → Limpiar datos → Crear embeddings → Almacenar en ChromaDB

2

Fase de Consulta

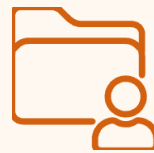
Pregunta usuario → Convertir a embedding → Buscar casos similares → LLM genera respuesta

Componentes Técnicos Clave



Embeddings

Representaciones numéricas que capturan significado del texto. Modelo: all-MiniLM-L6-v2 (384 dimensiones), ligero (23MB) y optimizado para español/inglés



ChromaDB

Base de datos vectorial que busca casos similares usando similitud matemática. Retorna los documentos más relevantes en milisegundos.



LLM

Recibe los textos relevantes desde ChromaDB. Los usa como contexto para escribir una respuesta en lenguaje natural. Solo se basa en lo recuperado

Preparación de Datos

01

Carga y Limpieza

Normalización de texto, manejo de valores nulos y estandarización de formatos
(para que los embeddings representaran bien el contenido)

02

Campos Adicionales

Banderas para casos de redes sociales y etiquetas como "acoso escolar"

03

Campo "Documento"

Combinación de columnas relevantes en texto estructurado
(para que el modelo pudiera capturar el contexto completo de cada caso)

04

Generación de Embeddings

329 casos convertidos a vectores numéricos

Estado Actual del Proyecto

✓ Completado

- Análisis del problema y definición de alcance
- Evaluación de alternativas técnicas
- Diseño de arquitectura RAG
- Limpieza y preparación de datos (329 casos)
- Generación de embeddings

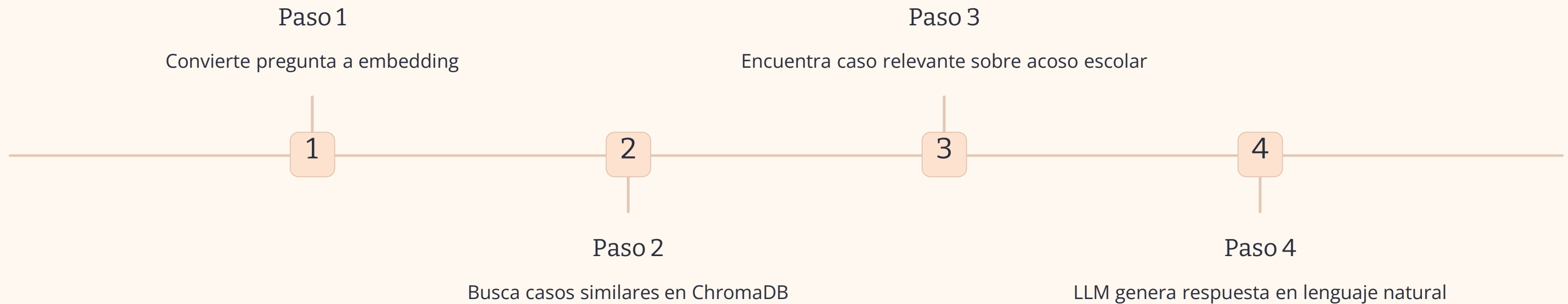
Pendiente

- Indexación en ChromaDB
- Implementación de sistema de retrieval
- Integración con LLM para respuestas
- Testing y validación

Cada pregunta del usuario se convierte en un vector, se busca en ChromaDB, y un modelo de lenguaje (LLM) genera la respuesta en texto natural.

Ejemplo de Uso Esperado

"¿Cuál fue la sentencia del caso sobre acoso escolar?"



Respuesta esperada: "En el caso de tutela T-XXX/2023, la sentencia determinó que..."

Transformando la Consulta Legal

De búsquedas manuales en Excel a respuestas automáticas inteligentes

329

Casos Procesados

Base de conocimiento

384

Dimensiones

Vectores semánticos
por caso

23MB

Modelo Ligero

Optimizado para
rendimiento

Para una prueba de concepto consideré importante demostrar viabilidad sin depender de infraestructura costosa.