Secure LLM Powering Semantic Search for a Multinational Development Bank

The Challenge

A multinational development bank with 48 member countries provides loans, grants, and technical assistance to its members for a wide range of development projects to reduce poverty and inequality in the region. The projects cover many topics, including infrastructure, education, healthcare, and environmental sustainability. Throughout the lifecycle of their projects, their subject matter experts (SMEs) create content in the form of blogs, articles, research papers, reports, and lessons learned.

All information related to a prospective project, including details about the client and performance and experience from comparable projects, must be evaluated before proceeding, especially during the proposal phase, to plan and price the project correctly. However, it has proven challenging to make all content findable within their current search experience entirely based on keyword searches. The content is typically stored in a nested file system, with each project having its own location and owner, resulting in information silos. Furthermore, with four official languages (English, Spanish, French, and Portuguese), surfacing content across languages and topics has proven inefficient and time-consuming for most of their users, whose query patterns range from single keywords to lengthy phrases.

The bank recognized it needed a better search experience to deliver its knowledge assets to its users at their point of need. The solution required implementing a unified search system that could seamlessly navigate through siloed content and break down language barriers. Additionally, it needed to provide efficient and accurate results for users with diverse query patterns, regardless of the length or complexity of their search queries.

The Solution

EK designed, developed, and deployed a hybrid search engine to address these issues. The search engine uses large language models (LLMs) to understand and interpret user requests in various languages. The search engine can deliver accurate results in all four languages using LLMs, eliminating users needing to switch between languages or complete separate queries. Furthermore, the hybrid approach combines traditional keyword-based search with natural language processing algorithms to provide a comprehensive and efficient search experience for customers with a wide range of query patterns.

Instead of relying solely on keywords, the search engine allows users to enter their queries more conversationally, allowing them to ask questions or provide specific details. Now, users can find the information they need quickly with an accuracy improvement in the search results, making the search engine easier to use and more intuitive for the user, resulting in an overall improvement in the quality of the user experience.

An open-source vector database that uses a fine-tuned language model to automatically vectorize and index content for quick retrieval powers the solution. The same language model is used at query time to vectorize the query and match it against the indexed content. Because of the use of the LLM to vectorize the content and the query, the search engine can return relevant results based on the query's semantic meaning. At the same time, a second process, keyword search, is initiated. The final result is a weighted score between keyword-based and vector-based search results based on several query attributes, such as length or terminology.

Secure LLM Powering Semantic Search for a Multinational Development Bank
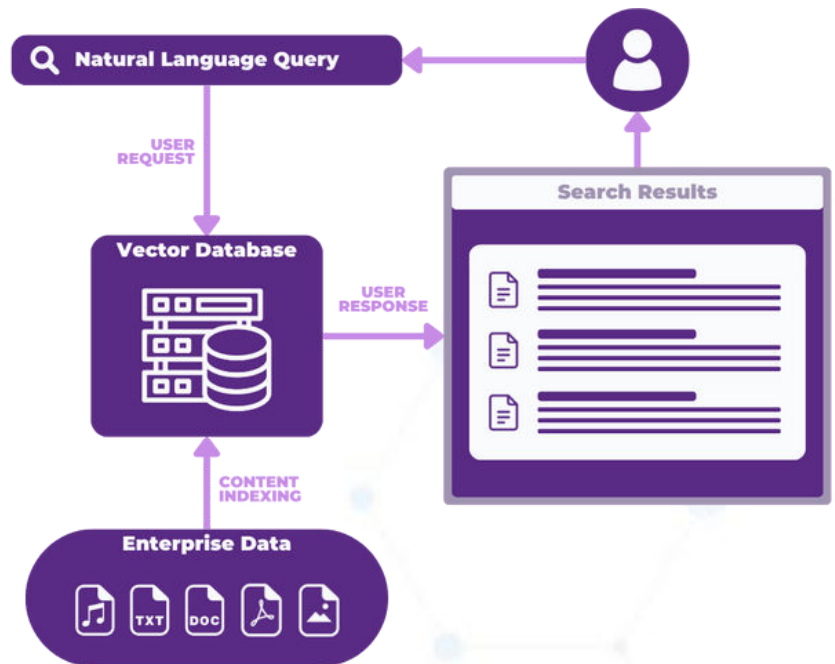
The EK Difference

EK formed a team of search experts, data engineers, and data scientists to plan and test the solution and provide a holistic infrastructure for the hybrid search feature. This team was able to evaluate different methods and components to ensure each of the elements was best-of-class for the end solution. The process involved testing several vector databases, pre-trained language models, and fine-tuning a multilingual language model on the bank's corpus to optimize the search algorithm.

Furthermore, the team researched user behavior and feedback extensively to understand users' specific needs and preferences. Combining their expertise with the insights gathered, EK developed a highly efficient and accurate hybrid search feature that provides relevant search results and understands the context and intent behind user queries. This hybrid search feature has dramatically improved the overall user experience and made it easier for the bank's users to find the information they require at the point of need.

The Results

The result is a powerful and user-friendly search tool that enhances productivity and eliminates language barriers for users across the bank's subsidiaries in member countries. This search tool has been instrumental in streamlining information and collaboration among users from different teams, allowing for seamless knowledge sharing and problem-solving. Furthermore, combining keyword and vector searches powered by LLMs ensures that search results are curated with AI augmentation but highly accurate and tailored to each user's specific needs, further improving their overall experience.

Natural Language Query

USER REQUEST

Vector Database

USER RESPONSE

Search Results

CONTENT INDEXING

Enterprise Data