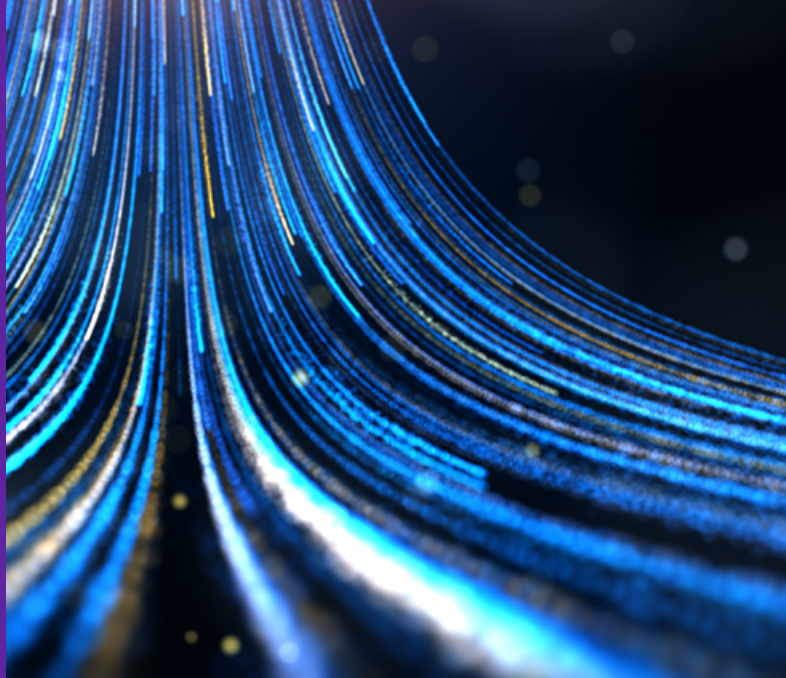## Luxoft
A DXC Technology Company

Generative AI accelerates data extraction from documents

Luxoft's experts deliver exceptional generative AI results for clients all over the world

Case study

Client:
A leading insurance firm

Industry:
Financial services

Project type:
Automating manual workflows

Technologies and tools:

Python

Web technologies

Vector DBs

Challenge

Retrieval augmented generation (RAG) combines two powerful components: Information retrieval and generative model capabilities. One groundbreaking application of generative AI-powered RAG was in the insurance sector. Our client's challenge was to transfer old insurance policies from a legacy system to a modern version. The massive volume of documents (dating back to 2005) from which data was traditionally extracted by hand, which took too long and occupied the valuable time of too many employees.

Our team expedited the process using RAG to develop an application, which enables analysts to interrogate the relevant data in a Q&A style. Generative AI (LLMs and word embeddings, in particular) increased efficiency and accuracy.

💡 Solution

- Documents (e.g., insurance rider documentation, product specifications and policy documents) are loaded and vectorized (transformed into a unique mathematical form) by an encoder, creating an efficient, machine-readable reference system. Even complex constructs like words and phrases are represented as vectors, ensuring that content isn't just stored, but context-aware in preparation for more intelligent data retrieval

- Business analysts interact with the documents via a Q&A UI (user interface), selecting a batch to query and then submitting questions. These questions undergo the same encoding process so that they exist within the same vector space as the stored documents

- A semantic search is performed on the encoded documents to return the relevant sections of each document intelligently. These sections and the original question are then sent to an LLM. The LLM processes the data and generates a direct, coherent answer to the question, referencing the source document

🏔️ Results

- The original target workflow for conversion has been so successful that the client asked to use the same technique in other areas (e.g., KYC [Know Your Customer])

- A full solution is currently in final development before being put into production
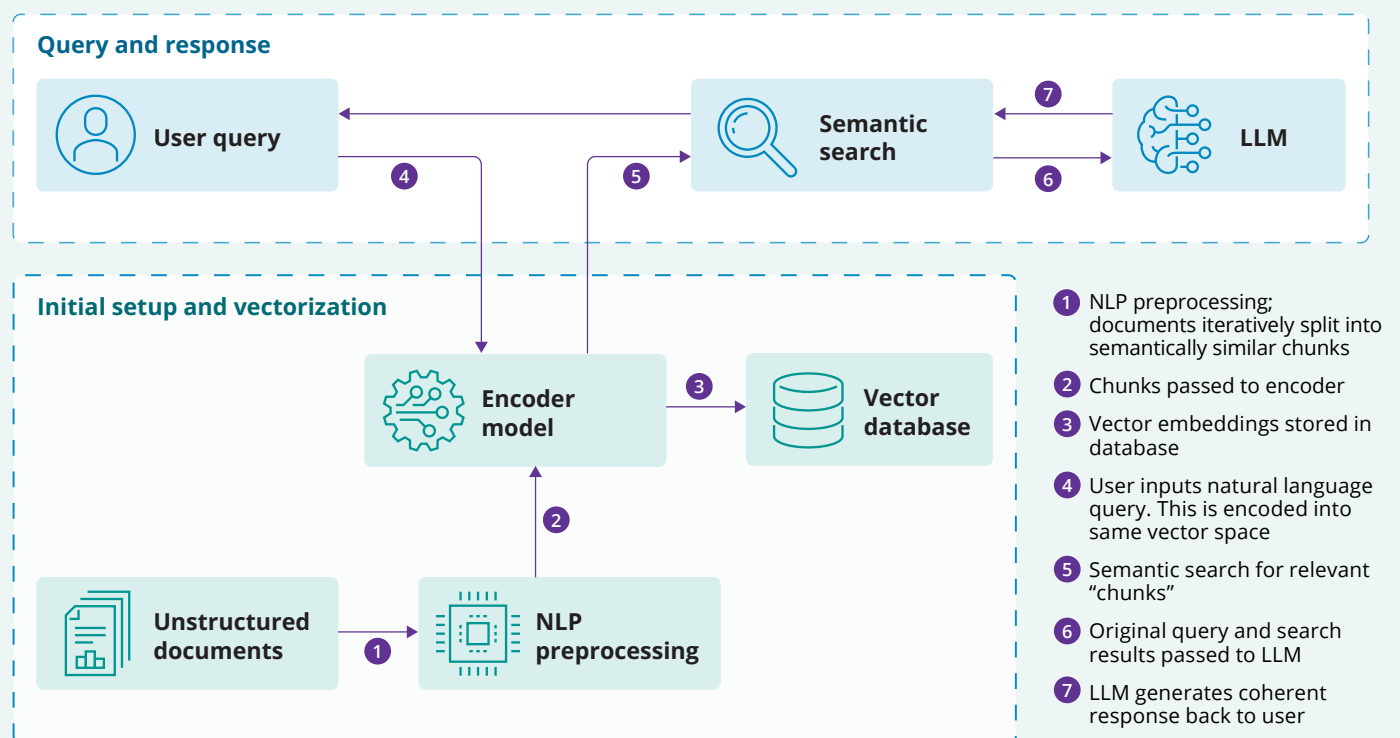
Empowered search via generative AI

The original workflow involved laboriously searching multiple documents and data sets. This provided a natural language interface that allows the user to "chat with their documents".

Users can perform more complex queries with multiple criteria and see concise results, in real time. The LLM can be hosted in-cloud or on-prem, depending on data privacy demands.

Query and response

User query

Semantic search

LLM

Initial setup and vectorization

Encoder model

Vector database

Unstructured documents

NLP preprocessing

1. NLP preprocessing; documents iteratively split into semantically similar chunks
2. Chunks passed to encoder
3. Vector embeddings stored in database
4. User inputs natural language query. This is encoded into same vector space
5. Semantic search for relevant "chunks"
6. Original query and search results passed to LLM
7. LLM generates coherent response back to user

Key benefits

Speed

What would have taken hundreds of employees several hours to complete has been reduced to a handful of staff and a few seconds. When searching the vector database, it takes just 38.3ms to return the closest matching section of a document for a single question (when searching over a million vectors).

User-friendly interaction

RAG allows users to request information in natural language and get the answer there and then, streamlining the user experience and boosting efficiency. No more navigating complex file systems or SharePoint databases.

Flexible querying

This technique enables the use of more complex, previously impossible queries. Queries like, "Show me all policies in the last 10 years with customers located in Birmingham, Liverpool and London with two or more cars" or "Compare the reinstatement interest across these products if the policyholder is 85 years old." Users can mix and match search types, combining various complex search criteria into a single query instead of having to make numerous separate searches.

Modularity

One of the greatest advantages of the solution is that depending on data privacy needs, businesses can either use a hosted LLM, or an on-prem model. This flexibility ensures that the solution adapts to many requirement types.

Looking to the future

As generative AI possibilities escalate, approaches like RAG will redefine industries, delivering insights and enhancing productivity in ways previously unimagined. This project typifies the transformative power of generative AI in real-world conditions and reinforces the potential of leveraging technological advancements for better business outcomes.

The future of data interrogation is super-efficient, intelligent and generative.

Talk to us

Want to discover how RAG can take your business to the next level? Visit the **Luxoft website** or **talk to one of our generative AI experts** today.

About Luxoft

Luxoft, a DXC Technology Company, is a trusted partner in global digital transformation and a leader in delivering competitive advantage in the software-defined world. We engineer and deliver innovative services and products that shape the future of industries by leveraging our extensive partnership network and deep industry-specific expertise.

For more information, please visit **luxoft.com**