

Tema:

Geração de relatórios sintéticos com LLMs e redução de texto baseado em busca semântica

Introdução

A Petrobras desenvolveu um sistema de respostas a perguntas sobre relatórios técnicos de inspeção em plataformas offshore de extração de óleo e gás, denominado SeSO (Semantic Search for Offshore Engineering). O projeto busca superar dois desafios principais: a escassez de dados reais, criando relatórios sintéticos com IA, e o limite de contexto dos LLMs na etapa de geração de respostas, investigando resumos que mantenham a qualidade das respostas.

Semantic Search for Offshore Engineering

O SeSO existente processa relatórios de inspeção da Petrobras transformando PDFs em XMLs para texto e tabelas, além de imagens para conteúdos visuais. Esses dados estruturados são divididos em trechos menores e indexados para a busca via Elasticsearch. Dada a pergunta de um usuário, o sistema recupera os trechos mais relevantes gera respostas com um LLM, garantindo agilidade e precisão na consulta a informações técnicas.

Material e Métodos

Criamos templates de relatórios baseados na estrutura dos existentes, e geramos novos relatórios com Chat GPT-4o. A partir deles, usamos Python com LangChain para expandir o dataset com versões reescritas, além de gerar as perguntas e respostas para testarmos a qualidade da geração. Quanto à busca semântica, utilizamos embeddings da OpenAI para identificar subtrechos irrelevantes e removê-los.

Conclusão

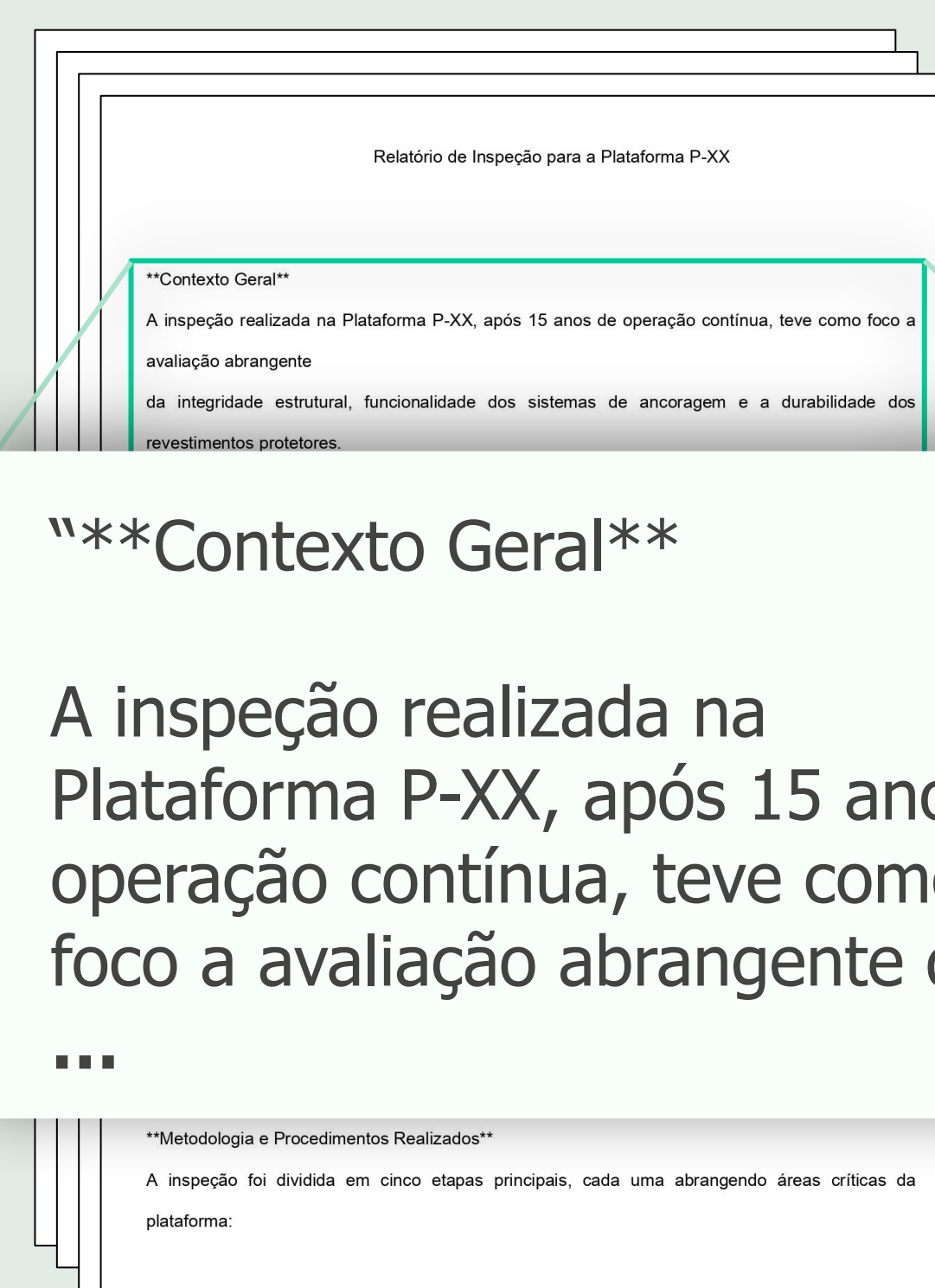
Os relatórios sintéticos gerados serão uma contribuição importante para o time que continua a desenvolver o SeSO, servindo como dataset para treinos e testes de modelos futuros. Além disso, a técnica de redução de texto mostrou-se útil como solução para superar as limitações de context window nos LLMs, permitindo o uso eficiente de trechos extensos sem perda de informações relevantes.

Visualização

Template de Relatório

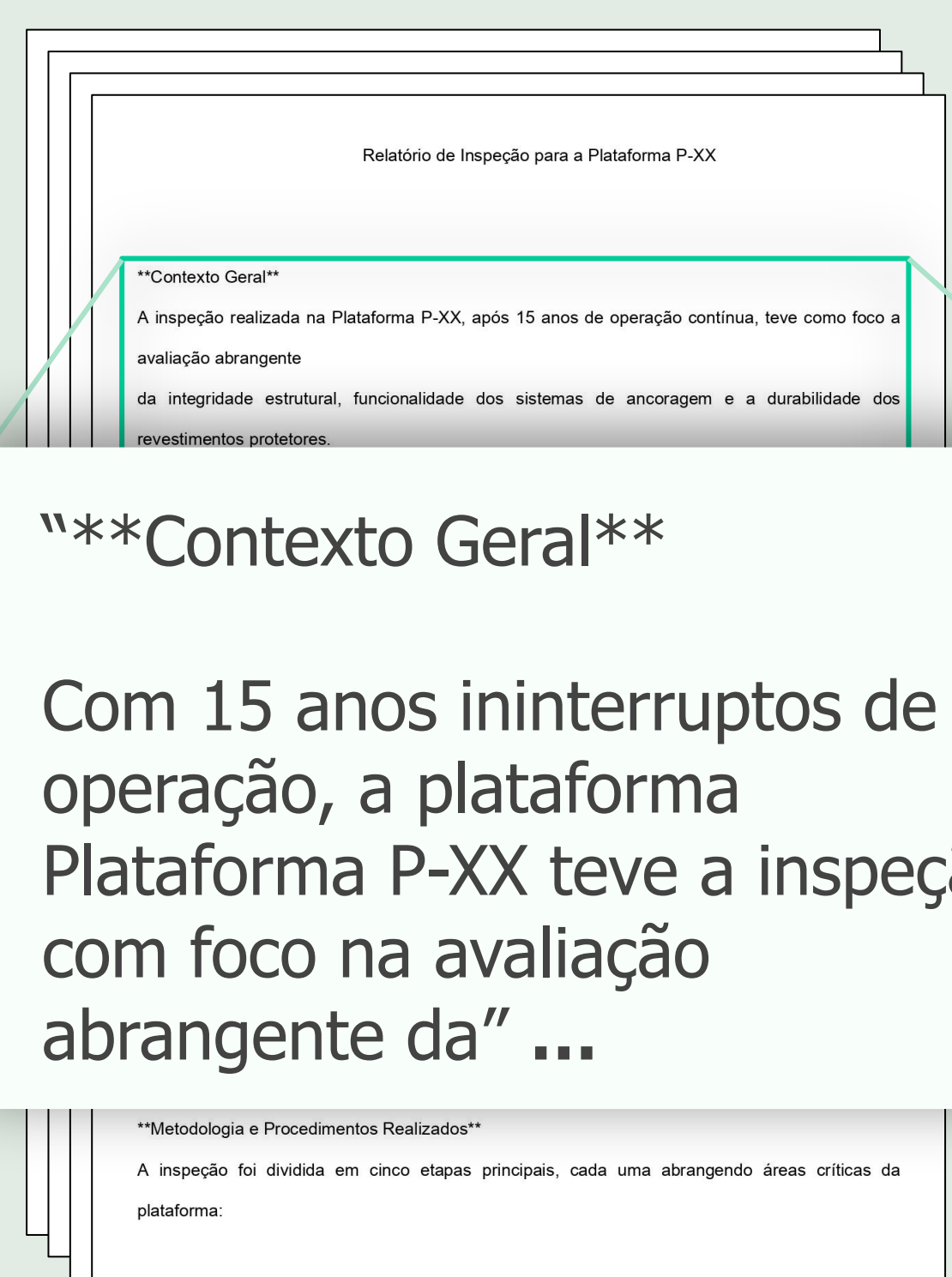
"gere um relatório como no template"

Relatório Sintético



para cada trecho
"reescreva o trecho"

Relatório Reescrito



para cada trecho
"crie 5 pares P&R"

P&Rs Geradas

P&R para trecho 1

P: "Há quanto tempo a plataforma P-XX está ativa?"; **R:** A plataforma está ativa há 15 anos

Trecho Reduzido

Eliminação de subtrechos dada a pergunta:

Há quanto tempo a plataforma P-XX está ativa?

~~Contexto Geral~~ Com 15 anos ininterruptos de operação ininterruptos de operação, a plataforma Plataforma P-XX teve a inspeção com foco na avaliação abrangente

"Com 15 anos ininterruptos de operação, a plataforma Plataforma P-XX"