Projeto de Formatura – 2024 – Press Release

PCS - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia de Computação

Tema: Synthetic text generation and retrieval text reduction based on embedding using natural language processing

---

Estudantes da Poli-USP desenvolvem solução de IA para otimização de sistemas de busca semântica e geração de relatórios sintéticos

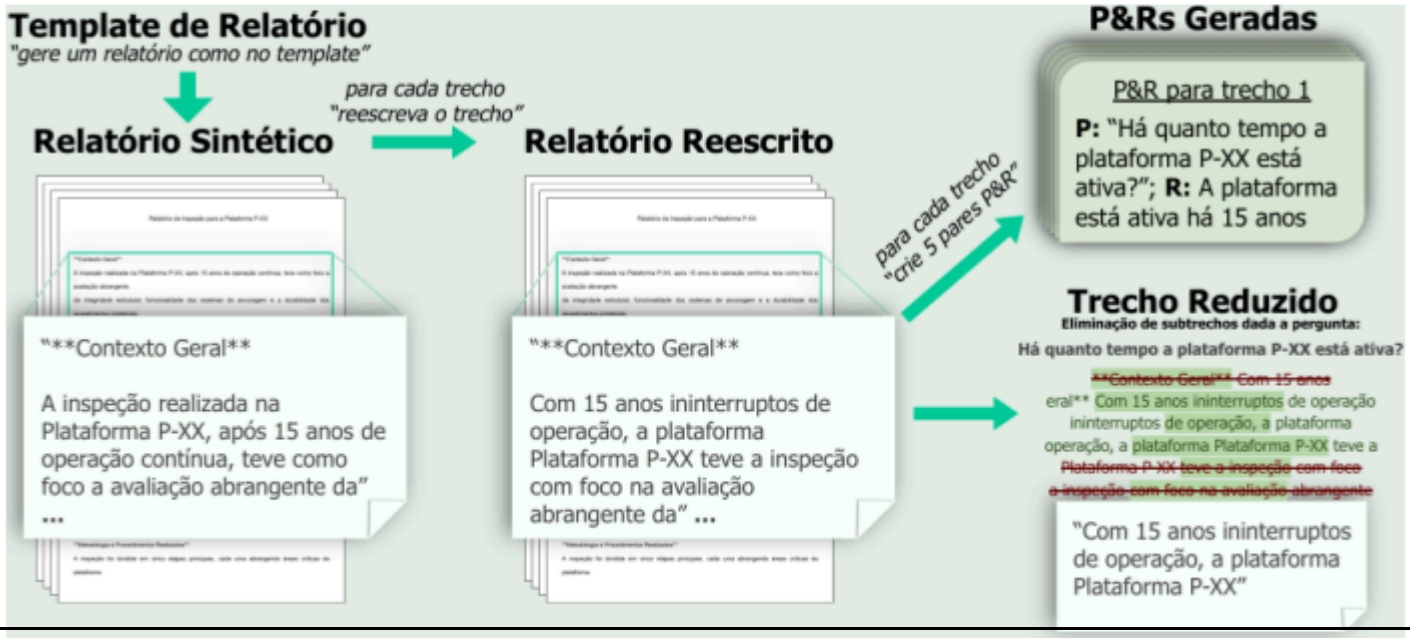
Os estudantes Angela Colas, Carlos Eduardo Jedwab e João Pedro Aras, do curso de Engenharia de Computação da Escola Politécnica da USP, Sob orientação do Professor Doutor Edson Gomi, apresentam o seu projeto que aborda a integração de inteligência artificial com sistemas de busca semântica para aprimorar a análise e a consulta de relatórios de inspeção.

A Petrobras utiliza o sistema SeSO (*Semantic Search for Offshore Engineering*) para processar relatórios de inspeção em plataformas offshore. No entanto, o sistema enfrenta dois grandes desafios: a escassez de dados reais para treinamento de modelos e as limitações de contexto dos *Large Language Models* (LLMs) na geração de respostas.

Para superar essas dificuldades, os estudantes desenvolveram um pipeline que combina:

1. **Geração de Relatórios Sintéticos:** Relatórios técnicos foram criados utilizando templates baseados em dados reais e gerados com o GPT-4o, garantindo alta fidelidade aos formatos originais
2. **Reescrita de Dados:** Relatórios e trechos textuais foram reestruturados para diversificar os dados disponíveis e melhorar o treinamento de sistemas de busca.
3. **Redução Textual Inteligente:** Por meio de transformações de word embeddings, subtrechos distantes do contexto da pergunta são considerados irrelevantes e são eliminados de grandes volumes de texto, otimizando o uso de LLMs sem perda de informações críticas.
4. **Geração de Perguntas e Respostas (P&Rs):** Foram criados pares de perguntas e respostas (P&Rs). Esses pares permitem testar e validar o desempenho do sistema, simulando interações reais de usuários.

Os relatórios sintéticos gerados ampliaram a base de dados disponível, contribuindo para o avanço do SeSO. A técnica de redução textual mostrou-se essencial para superar as limitações de tamanho de contexto das LLMs, permitindo consultas mais precisas e eficientes. Além disso, o uso de tecnologias como Python, LangChain e OpenAI Embeddings garantiu um processamento ágil e confiável.



Integrantes: Angela Colas, Carlos Eduardo Jedwab, João Pedro Aras

Professor(a) Orientador(a): Prof. Dr. Edson Satoshi Gomi