

Breve Descrição

A aplicação é uma ferramenta utilitária cujo uso principal consiste em resumir artigos fornecidos pelo usuário usando LLMs, gerando evidence briefings. Evidence briefings são resumos objetivos que destacam descobertas e implicações de evidências para públicos específicos.

O modelo utilizado no projeto é o gpt-4-mini integrado com RAG (Retrieval-Augmented Generation) alimentado por uma base de evidence briefings para referência. A principal função do programa é transformar artigos acadêmicos em resumos mais acessíveis e práticos, especialmente voltados para profissionais da indústria.

Funcionalidades específicas incluem:

- Processamento e sumarização de artigos acadêmicos em formato amigável para praticantes do mercado;
- Busca e recuperação de informações relevantes nos documentos para contextualizar os briefings;

Siga as instruções contidas no arquivo README.md contido na raiz do projeto para instruções de execução.

Visão de projeto

Cenário Positivo 1

Joana, uma gerente de projetos na indústria farmacêutica, recebe um artigo acadêmico com descobertas recentes sobre métodos de produção de medicamentos. Ela utiliza a aplicação para carregar o documento e gera um evidence briefing em poucos minutos. O resumo inclui os principais achados e suas implicações práticas. Joana rapidamente entende como aplicar os insights no planejamento de novos projetos e compartilha o briefing com sua equipe, economizando horas de leitura e interpretação.

Cenário Positivo 2

Marcos, um estudante de engenharia de software, está trabalhando em um projeto de pesquisa sobre técnicas de machine learning aplicadas na indústria. Ele utiliza o programa para gerar evidence briefings de artigos desejados relacionados ao tema. Quando precisa revisar sua pesquisa, Marcos consulta os briefings gerados para identificar rapidamente os principais achados e limitações dos artigos, ajudando-o a estruturar melhor sua análise e organizar as fontes de maneira eficiente.

Cenário Negativo 1

Ana, uma profissional de marketing, tenta carregar um PDF que contém apenas imagens escaneadas do texto. O programa processa o arquivo normalmente, mas, ao tentar gerar o evidence briefing, a aplicação responde com:

“Sure! Please provide the details of the paper you'd like me to summarize for the evidence briefing.”

Ana percebe que o seu arquivo não possui texto selecionável e entende que precisará fornecer outro arquivo.

Cenário Negativo 2

Carlos, um pesquisador na área de biotecnologia, carrega um artigo detalhado sobre novas técnicas de edição genética. Apesar de o programa gerar um evidence briefing, Carlos percebe que a sumarização omitiu informações importantes sobre os parâmetros específicos utilizados nos experimentos descritos no artigo. Ele entende que precisará consultar o documento original para obter os detalhes técnicos necessários para sua análise aprofundada.

Documentação Técnica do Projeto

O código conta com documentação docstring para todas as classes e métodos, seguindo padrões de estilo, para facilitar o entendimento, manutenção e extensão do código.

Especificação de requisitos

Requisitos Funcionais

RF1 : O software deve permitir o upload de arquivos .pdf para gerar evidence briefings.

RF2 : O software deve listar os arquivos disponíveis para processamento.

RF3 : O software deve processar e retornar uma evidence briefing, em formato textual, para um dos arquivos selecionados.

Requisitos Não Funcionais

RNF1 : O software deverá gerar um dado evidence briefing em no máximo 2 minutos.

Arquitetura Geral

O sistema é dividido em componentes principais que colaboram para processar documentos e gerar evidence briefings. Os módulos presentes na solução estão detalhados abaixo:

View:

- Responsável por permitir que o usuário carregue documentos, selecione arquivos disponíveis e visualize os evidence briefings gerados.

Controller:

- Módulo centralizador que gerencia o passo a passo de uma execução, inicializando os outros módulos e os chamando quando necessário

Data Loader:

- Módulo responsável por carregar a base de pdfs para a criação do banco vetorial e pelo carregamento de um único pdf.

Prompter:

- Módulo responsável pela fabricação dos prompts necessários para a aplicação, recebendo tanto os trechos relevantes ao briefing quanto o documento do usuário a ser resumido.

Retriever:

- Módulo responsável pelo vector database. Contém o banco com os briefings vetorizados e gerencia a busca de trechos de texto relevantes.

Model:

- Responsável por chamadas diretas ao modelo.

Bibliotecas utilizadas

- Streamlit para renderização da interface com o usuário.
- Chroma, biblioteca responsável pelo banco de dados vetorial que alimenta o RAG.
- Langchain, possui classes auxiliares como:
 - PyPDFLoader
 - ChatOpenAI, biblioteca usada para comunicação com o modelo gpt-4o-mini da OpenAi.
 - SentenceTransformerEmbeddings, responsável por vetorizar dados para serem armazenados no ChromaDB.


Manual do Usuário




A tela acima é o estado inicial da aplicação. Nesse ponto, o usuário deve clicar ou arrastar um arquivo no botão “Browse Files” a ser resumido.

Gerenciamento de Arquivos

Upload do arquivo

 Drag and drop file here
Limit 200MB per file • PDF

Browse files

 2308.10620v6.pdf 2.4MB

×

Arquivo 2308.10620v6.pdf salvo com sucesso!

Selecione um arquivo para processar

Arquivos disponíveis:

2308.10620v6.pdf

Categoria do Artigo (opcional):

Processar arquivo selecionado


Briefing Gerado

Nenhum resultado disponível.


Após subir qualquer arquivo pdf, o usuário deve escolher no menu dropdown o arquivo desejado e, opcionalmente, adicionar uma categoria para o seu artigo. Essa categoria ajuda a aplicação a direcionar a busca nas evidence briefings de referência. Após essa etapa, o botão “Processar arquivo selecionado”.

Gerenciamento de Arquivos

Upload do arquivo

 Drag and drop file here
Limit 200MB per file • PDF

Browse files

 2308.10620v6.pdf 2.4MB

×

Arquivo 2308.10620v6.pdf salvo com sucesso!

Selecione um arquivo para processar

Arquivos disponíveis:

2308.10620v6.pdf

Categoria do Artigo (opcional):

Systematic Literature Review

Processar arquivo selecionado

Briefing Gerado

Resultado:

Evidence Briefing: Large Language Models for Software Engineering - A Systematic Literature Review

Intro:
This briefing is based on the systematic literature review conducted by Xinyi Hou et al., which explores the impact of Large Language Models (LLMs) in the field of Software Engineering (SE).

What is Systematic Reviews?
Systematic reviews are comprehensive, structured assessments of existing research on a specific topic, aimed at providing a clear understanding of the current state of knowledge. They involve systematic searching, selection, and synthesis of studies to answer defined research questions.

For more information about systematic reviews, visit: [Systematic Reviews](https://cin.ufpe.br/eseq/systematic-reviews).

What is included in this briefing?

1. **“Main Findings”**: The briefing summarizes key findings from the original systematic review, highlighting how LLMs can optimize software engineering processes and outcomes.
2. **“Evidence Characteristics”**: It includes a brief description of the original systematic review and the studies it analyzed, focusing on the categorization of LLMs, data handling techniques, tuning methods, evaluation metrics, and the specific SE tasks addressed.

O briefing gerado se encontra na caixa de texto na coluna da direita.