

Arquitetura do Projeto de Processamento de Dados da Câmara dos Deputados

1. Coleta de Dados (dataprep.py)

- API da Câmara dos Deputados
- Arquivo .env para chave Gemini
- Processadores de dados: Sumarização, análise de sentimentos, marcação de palavras-chave
- Vetorização para bases de referência

2. Armazenamento de Dados

- Parquet para dados tabulares
- JSON para sumários e metadados
- FAISS Index para buscas vetoriais
- YAML para configurações

3. Dashboard (dashboard.py)

- Abas de Visualização: Resumos, Análise de Sentimentos, Pesquisa Vetorial
- Funcionalidades interativas e uso de LLMs

4. Modelos de Linguagem de Larga Escala (LLMs)

- Sumarização e geração de respostas