



FROM NATURAL LANGUAGE TO SQL

Um framework leve em Python usando
LLMs para consultas em bancos relacionais

REDES NEURAIS E DEEP LEARNING

Prof. Adrião Duarte Dória Neto

Efrain Marcelo Pulgar Pantaleon

Willian Moura Gondim de Freitas

Pietro Augusto de Albuquerque Lira e Silva



FROM NATURAL LANGUAGE TO SQL

Um framework leve em Python usando LLMs para consultas em bancos relacionais

2

OBJETIVO DO TRABALHO

- Criar um framework leve baseado em LLMs (GPT-4, LLaMA 3).
- Traduzir linguagem natural para SQL em consultas zero-shot.
- Garantir segurança, precisão e integração com banco de dados.

FROM NATURAL LANGUAGE TO SQL

Um framework leve em Python usando LLMs para consultas em bancos relacionais

MOTIVAÇÃO

- 87% dos usuários não técnicos têm dificuldade com SQL.
- Geração de relatórios pode levar até 5 dias.
- Soluções tradicionais NL2SQL requerem treinamento extenso.
- LLMs oferecem alternativas promissoras, mas com desafios.

FROM NATURAL LANGUAGE TO SQL

Um framework leve em Python usando LLMs para consultas em bancos relacionais

CONTRIBUIÇÕES PRINCIPAIS

- Pipeline de 5 etapas com geração de SQL zero-shot.
- Avaliação quantitativa (92% consultas simples / 68% joins).
- Código aberto com adaptadores para PostgreSQL/SQLite.

INTEGRAÇÃO COM O LLM

- Ingestão de esquema com:
- Tabelas, colunas e tipos
- Chaves estrangeiras (grafo)
- Dados exemplo
- Prompts otimizados com diretrizes de segurança.

FROM NATURAL LANGUAGE TO SQL

Um framework leve em Python usando LLMs para consultas em bancos relacionais

ARQUITETURA DO SISTEMA

- LLM Agent: Recebe schema e gera SQL.
- Validação em Python: Regex, parsing e segurança.
- Execução + Pós-processamento: Anonimização e timeout.

PIPELINE DE EXECUÇÃO

- Sanitização de queries
- Pool de conexões
- Timeout de 5s
- Pós-processamento dos resultados
- Execução média: < 2 segundos para até 3 joins.

FROM NATURAL LANGUAGE TO SQL

Um framework leve em Python usando LLMs para consultas em bancos relacionais

BENEFÍCIOS EM PERFORMANCE

- Precisão:
- 92% de acurácia em consultas simples.
- 68% de acurácia em consultas com múltiplos JOINS.
- Eficiência:
- Redução do tempo de geração de relatórios de 5 dias para menos de 15 minutos.
- Até 38 queries por minuto com otimizações (paralelismo, caching e rejeição precoce).

FROM NATURAL LANGUAGE TO SQL

Um framework leve em Python usando LLMs para consultas em bancos relacionais

CÓDIGO ABERTO DO SISTEMA

- Disponível para uso e experimentação com adaptadores para:
- PostgreSQL
- SQLite

APLICAÇÕES PRÁTICAS

- Saúde: relatórios de pacientes em tempo real.
- E-commerce: ajustes de preços dinâmicos com base em dados históricos.
- Educação e Governo: análises estatísticas com linguagem natural.
- Empresas com grande volume de dados e poucos especialistas em banco.

FROM NATURAL LANGUAGE TO SQL

Um framework leve em Python usando LLMs para consultas em bancos relacionais

DESAFIOS SUPERADOS

- Precisão semântica: Redução de erros em 41%
- Escalabilidade de esquema: Clusterização de tabelas
- Eficiência: Até 38 queries/min (caching + paralelismo)

RESULTADOS

- Base de dados com 10K registros
- 92% de acurácia (consultas simples)
- 68% de acurácia (joins complexos)
- Redução de tempo de geração: 5 dias → 15 minutos

FROM NATURAL LANGUAGE TO SQL

Um framework leve em Python usando LLMs para consultas em bancos relacionais

CONCLUSÃO

- O framework mostra que LLMs podem gerar SQL a partir de linguagem natural com alta precisão, sem necessidade de treinamento prévio.
- Reduz tempo e custo na geração de relatórios, facilitando o acesso a dados por usuários não técnicos.
- A solução é leve, segura e adaptável a qualquer banco relacional (PostgreSQL, SQLite).
- Combina engenharia de prompts, validação e execução otimizada, garantindo agilidade e confiabilidade.
- É útil em áreas como saúde, comércio, educação e gestão pública, onde decisões rápidas são essenciais.