



FIOCRUZ

Concurso Público Fiocruz 2023

Tecnologista em Saúde Pública

Prova Discursiva

TE56

Cientista de Dados em Saúde

Espelho de Resposta

Pontuação da Questão Discursiva conforme Anexo II do Edital nº 2, de acordo com a Unidade detentora da vaga.

Espera-se que o candidato, no desenvolvimento do tema, tenha feito considerações técnicas adequadas sobre os seguintes pontos:

1 – Inicialmente, o Llamalndex fornece ao modelo de linguagem o esquema do banco de dados. O modelo de linguagem, ao receber o esquema do banco de dados como parte do prompt, usa essa informação para entender a estrutura. Neste momento, nenhum dado é apresentado ao modelo. Em seguida, ele mapeia a intenção da pergunta para a lógica SQL correspondente, gerando uma consulta SQL que busca a resposta desejada no banco de dados, considerando as relações, atributos e restrições definidos no esquema.

2 – *SELECT SUM(f.valor) AS total_valor
FROM financiamento f
JOIN projeto p ON f.projeto_id = p.projeto_id
JOIN pesquisador pe ON p.projeto_id = pe.projeto_id
WHERE EXTRACT(YEAR FROM f.data_inicio) = 2023
GROUP BY f.projeto_id
HAVING COUNT(DISTINCT pe.pesquisador_id) >= 2*

3 – Não, os dados do banco de dados não são conhecidos pelo modelo de linguagem no momento da geração do código SQL na técnica Text-to-SQL. O modelo utiliza informações sobre o esquema do banco de dados, como nomes de tabelas e colunas, para entender a estrutura dos dados. No entanto, o modelo não tem acesso direto ou conhecimento dos dados contidos no banco. Ele gera consultas SQL baseando-se na pergunta em linguagem natural e no esquema fornecido, mas sem conhecimento prévio dos dados específicos armazenados.

Por exemplo, ao utilizar modelos Codex da OpenAI para tradução de texto para SQL, as informações do esquema do banco de dados são enviadas junto com a instrução em linguagem natural para o modelo. Este processo ajuda o modelo a gerar o SQL solicitado de forma mais precisa, considerando a estrutura específica do banco de dados em questão. Assim, embora o modelo de linguagem não

"conheça" os dados do banco de dados inherentemente, ele utiliza as informações do esquema fornecidas para entender o contexto e gerar consultas SQL corretas.

4 – O LlamaIndex executa a consulta no banco de dados e os resultados obtidos são repassados ao modelo de linguagem, junto com a pergunta e quaisquer instruções necessárias no *prompt*. Esses resultados são então processados e formatados de maneira a serem compreensíveis e informativos para o usuário. A apresentação da resposta final pode variar dependendo da interface do usuário, podendo ser exibida em forma de texto, tabelas, gráficos ou até mesmo como parte de uma conversa interativa em um *chatbot*.