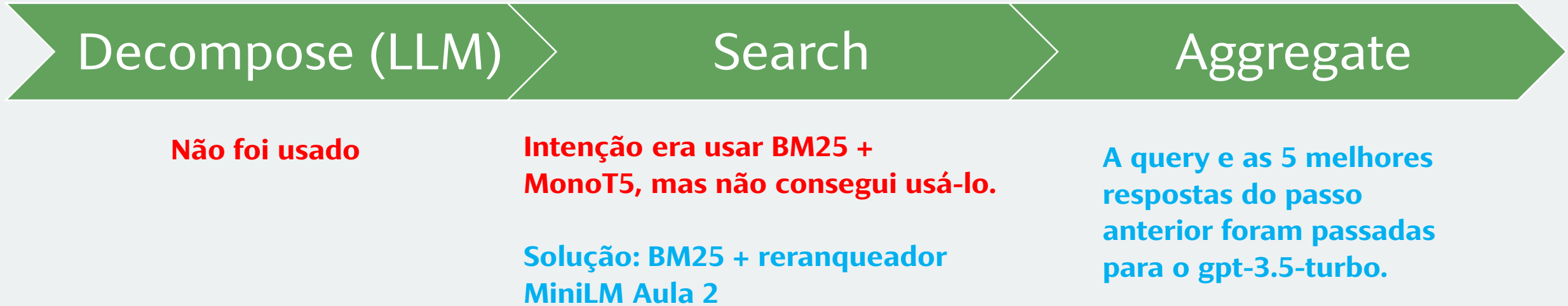


Notebook – Visconde

Leandro Carísio

Problemas e soluções no desenvolvimento



Considerações feitas

- Usei as 100 primeiras perguntas do arquivo `iirc_test.json`.
- Considerei apenas os artigos de `context_articles.json` que eram referenciados em `iirc_test.json` (~ 7000 artigos, ~500.000 janelas de sentenças).
- Few-shot: 2 exemplos fixos para cada chamada. Em cada exemplo, 4 documentos retornados.
- O resultado final foi processado para pegar a string depois de "Answer: ". Além disso, resultados como "unknown", "it is not possible...", "no information..." foram uniformizados para "none".

Prompt usado com problema

Pedi pra dar evidências e respostas diretas:

*You are a helpful reader that reads a list of documents and use them as evidence to answer a question. **You only give direct answers.***

Em vez de usar a role assistant, usei assitant (com typo). **Na prática era um zero-shot?**

F1: 0.33 / EM: 0.30 (81 questões)

F1: 0.21 / EM: 0.17 (removendo respostas none, 58 questões)

Prompt final

You are a helpful reader that reads a list of documents and use them as evidence to answer a question. You provide detailed explanations and concise answers..

F1: 0.377 / EM: 0.31 (100 questões)

F1: 0.257 / EM: 0.159 (removendo respostas none, 69 questões)

Dúvida básica

Nesse exercício usamos um subconjunto de `context_articles`, considerando apenas os documentos apontados no `test_set`.

Não deveríamos ter usado todo o `context_article` em vez de menos de 15% dele?

Obrigado

Leandro Carísio
carisio@gmail.com