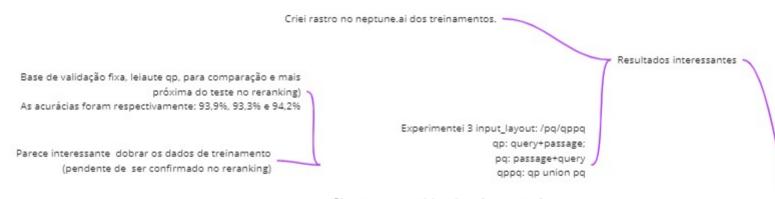


Código para Aula 3

IA368DD_2023S1: Deep Learning aplicado a Sistemas de Buscas Marcus Vinícius Borela de Castro

Classificação de Texto e Reranqueadores



Obs.: qppq com o dobro de registros no treinamento

Usei asserts no meio do código

Validei os códigos das funções com o chatgpt

Criei dados fictícios para testar a classe do dataset

Técnicas usadas para buscar a correção

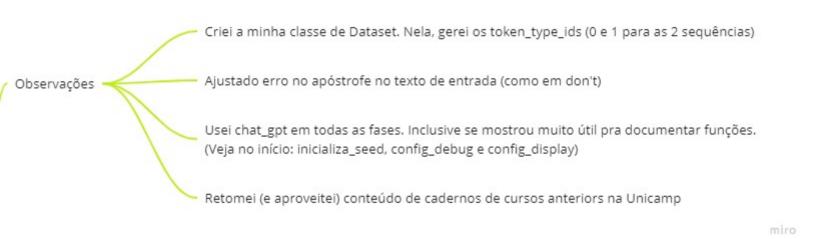


No dataset: Precisamos gerar ou o modelo gera para nós quando não passamos?

Dúvida: Não salvei o tokenizador pois ele não foi alterado. Precisa?

miro

Dúvidas



Ressalvas

Na fase do reranking, o ndcg_cut_10 deu baixo 0.0441.

Suspeita de causa: não ter usado torch.save(model) ao invés de model.save_pretrained(). Segundo ChatGPT, há uma diferença: model.save_pretrained() salva os pesos do modelo, bem como todos os outros artefatos necessários para carregar e inicializar o modelo em uma instância específica de um modelo do Transformers, enquanto torch.save(model) salva apenas os pesos do modelo.

Obs.: Fiz o ajuste no caderno e deixei treinando para ver a diferença.