

Leitura dos artigos

Document Expansion by Query Prediction
(Nogueira et al)

From doc2query to docTTTTTquery
(Nogueira et al)

Conceitos importantes

“***Vocabulary mismatch***”: Sinônimos > usuários usam termos de busca diferentes do que estão no documento.

“***Query expansion***”: Alterar a query do usuário colocando novos termos relacionados aquela query. É feito no momento da busca

Contribuições do artigo

“*Document expansion*”: Expande o documento com novas possíveis queries. É feito antes da indexação



As queries geradas possuem palavras copiadas do documento e palavras que não existiam nos documentos. As primeiras alteram a frequência das palavras e ajudam na reordenação do documento. As últimas ajudam a encontrar novos documentos.

Dúvida

Para o treinamento do modelo foi usado um dataset de query vs documentos relevantes.

Poderia ser usado também um dataset de query vs documentos não relevantes para gerar um segundo modelo? Esse segundo modelo popularia um campo separado, que depois poderia ser usado para diminuir o score pelo BM25 (por exemplo, o score seria dado pelo score do documento menos x% do score calculado usando apenas esse campo).

Obrigado

Leandro Carísio
carisio@gmail.com