

# Document Expansion by Query Prediction

...

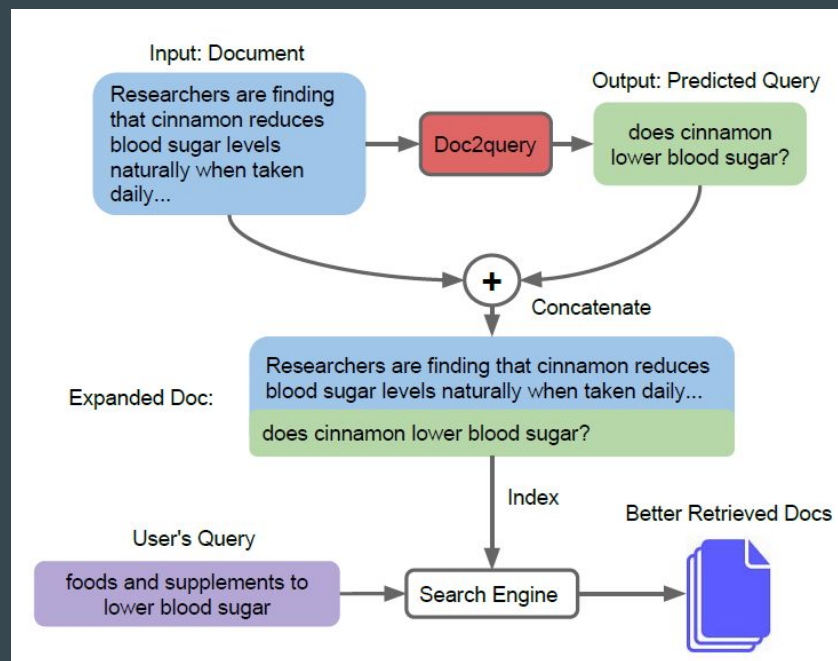
13 de abril de 2023

# Conceitos Importantes

- “**vocabulary mismatch**” - “**incompatibilidade de vocabulário**”, é um dos desafios centrais na recuperação de informações. É o problema em que os usuários usam termos de consulta que diferem daqueles usados em documentos relevantes.
- Aplicado na antes da indexação.

# Contribuições do Artigo

- Mostrou como enriquecer os documentos para a indexação utilizando modelo seq to seq, concatenando aos documentos as queries geradas.



# Contribuições do Artigo

- A técnica se mostra eficaz pois não altera o pipeline de busca
- Inferência rápida pois é possível utilizar somente o BM25 com bons resultados
- Se fizer reranqueamento com BERT melhora ainda mais os resultados

	TREC-CAR MAP Test	MS MARCO MRR@10 Test	Dev	Retrieval Time ms/query
Single Duet v2 (Mitra and Craswell, 2019b)	-	24.5	24.3	650*
Co-PACRR♠ (MacAvaney et al., 2017)	14.8	-	-	-
BM25	15.3	18.6	18.4	50
BM25 + RM3	12.7	-	16.7	250
BM25 + Doc2query (Ours)	18.3	21.8	21.5	90
BM25 + Doc2query + RM3 (Ours)	15.5	-	20.0	350
BM25 + BERT (Nogueira and Cho, 2019)	34.8	35.9	36.5	3400†
BM25 + Doc2query + BERT (Ours)	<b>36.5</b>	<b>36.8</b>	<b>37.5</b>	3500†

# Obrigado

Manoel Veríssimo

---