Prefix-Oriented Ranking Optimization

Query: what happen when you use regular soap to wash your hair

Pos. DocID:

179-150-49-111-100-143-85-252 > 179-72-101-26-95-100-132-220

Neg. DocID:

Pos. prefixes:

179-150

179-150-49

⋮

179-150-49-111-100-143-85

Neg. prefixes:

> 179-72

> 179-72-101

⋮

> 179-72-101-26-95-100-132

weight: $\alpha_i = \frac{1}{Z}(1 - \frac{\beta}{i})$

progressive training

Relevance-Based DocID Initialization

(a) Generative retrieve model:

dot product

Enc. → Dec.

query   <s>

Enc. → Dec.

query   <s>

Generative Retrieval Model M: Enc. → Dec.

(b) DocID construction via RQ

≈ + + ... +

Doc → DocID

Optimization Details

Relevance-Based DocID Initialization

M^0

→ Seq2Seq Pretraining

M^1

→ Rank-Oriented Fine-tuning

Initial Fine-tuning

M^2

→ Prefix-Oriented Ranking Optimization

M^3

→ Self-Negative Fine-tuning

M^4