

Summary: Ranking based on Query Likelihood

$$q = w_1 w_2 \dots w_n \quad p(q | d) = p(w_1 | d) \times \dots \times p(w_n | d)$$

$$f(q, d) = \log p(q | d) = \sum_{i=1}^n \log p(w_i | d) = \sum_{w \in V} c(w, q) \log p(w | d)$$

Document language model

Retrieval problem → Estimation of $p(w_i | d)$

Different estimation methods → different ranking functions