

Retrieval as LM Estimation

- Document ranking based on *query likelihood*

$$\log p(q | d) = \sum_{i=1}^m \log p(q_i | d) = \sum_{i=1}^{|V|} c(w_i, q) \log p(w_i | d)$$

where, $q = q_1 q_2 \dots q_m$

↑
Document language model

- Retrieval problem \approx Estimation of $p(w_i|d)$
- Smoothing is an important issue, and distinguishes different approaches