

# Benefit of Rewriting

- Better understanding of the ranking function
  - Smoothing with  $p(w|C) \rightarrow$  TF-IDF weighting + length norm.

**TF weighting**

**Doc length normalization**

**matched query terms**

**IDF weighting**

**Ignore for ranking**

$$\log p(q | d) = \sum_{\substack{w_i \in d \\ w_i \in q}} \left[ \log \frac{p_{\text{Seen}}(w_i | d)}{\alpha_d p(w_i | C)} \right] + n \log \alpha_d + \boxed{\sum_{i=1}^n \log p(w_i | C)}$$

- Enable efficient computation