Benefit of Rewriting

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

TF weighting

Doc length normalization

$$\log p(q \mid d) = \sum_{\substack{w_i \in d \\ w_i \in q}} c(w,q)[\log \frac{p_{Seen}(w_i \mid d)}{\alpha_d \, p(w_i \mid C)}] + n \log \alpha_d + \boxed{\sum_{i=1}^{n} \log p(w_i \mid C)}$$

matched query terms

IDF weighting

Ignore for ranking

- Enable efficient computation