

Pyserini: Dirichlet Smoothing Query Likelihood

$M=2000$

$$P(q|d) = \prod_{t \in q} \frac{tf(t,d) + M \cdot P(t|C)}{|d| + M}$$

Actual implementation:  $S_{\text{Dir}}(q, d) = \sum_{t \in q} q \cdot tf(t, q) \cdot \log \left( \frac{1 + \frac{tf(t, d)}{M \cdot \frac{tf(t, C)}{|C|}}}{\frac{|d| + M}{M}} \right)$

Evolved Step 182:

$$S(q, d) = \sum_{t \in q} w(t) \cdot \tilde{S}(t, d) + \sum_{t \in q} m(t, d) + AND(q, d) + LP(d)$$

$$w(t) = (q \cdot tf(t) \cdot r(t))^{0.6}$$

$$r(t) = 1 + 0.9 \cdot \min(\max(\log \frac{P_{df}(t)}{P_c(t)}, 0), 2.5)$$

$$S_{\text{base}}(t, d) = g(t) \cdot \log \left( \frac{1 + \frac{tf(t, d)}{1750 \cdot P_c(t)}}{\frac{|d| + 1750}{1750}} \right)$$

$$P_c(t) = 0.97 (0.9 \cdot P_r(t) + 0.1 \cdot P_{df}(t)) + \frac{0.03}{|V|}$$

$$P_r(t) = \frac{P(t|C)^{0.85}}{\sum_{t' \in V} P(t'|C)^{0.85}}$$

$$P(t|C) = \frac{tf(t, C)}{|C|}$$

$$\tilde{S}(t, d) = \begin{cases} S_{\text{base}}(t, d) & \text{if } S_{\text{base}}(t, d) \geq 0 \\ 0.12 \cdot S_{\text{base}}(t, d) & \text{otherwise} \end{cases}$$

$$g(t) = 1 + 0.45 \cdot \text{clip} \left( \log \left( \frac{P_{df}(t)}{P_c(t)} \right), -2.5, 2.5 \right)$$

$$\beta(t) = 1 - 0.3 \cdot (1 - IDF_{01}(t))$$

$$IDF_{01}(t) = \frac{\log \frac{N+1}{df(t)+1}}{\max_{t' \in V} \log \left( \frac{N+1}{df(t')+1} \right)}$$

$$m(t, d) = \mathbb{1}[tf(t, d) = 0] \cdot 0.07 \cdot w(t) \cdot \log \left( \frac{1750 \cdot P_c(t)}{|d| + 1750} \right)$$

$$AND(q, d) = 0.14 \cdot \frac{1}{|q|} \sum_{t \in q} \tanh \left( \frac{w(t) \cdot \max(\tilde{S}(t, d), 0)}{3} \right)$$

$$LP(d) = -0.06 \cdot (\log(|d|) - \log(\text{avg}|d|))^2$$