

$$\text{Pyserini BM25: } \text{score}(q, d) = \sum_{t \in q} \log\left(1 + \frac{N - df(t) + 0.5}{df(t) + 0.5}\right) \cdot \left( \frac{tf(t, d) \cdot (k_1 + 1)}{tf(t, d) + k_1(1 - b + b \cdot \frac{|d|}{avgdl})} \right)$$

$k_1 = 0.9$   
 $b = 0.4$

Evolved Step 297:

$$\text{score}(q, d) = R(q^{\text{base}}, d^{\text{base}}) + w_{\text{pfx}} R(q^{\text{pfx}}, d^{\text{pfx}}) + w_{\text{bi}} R(q^{\text{bi}}, d^{\text{bi}}) + w_{\text{mic}} G(q^{\text{mic}}) R(q^{\text{mic}}, d^{\text{mic}})$$

$G(q) = \sigma\left(\frac{\overline{\text{IDF}}(q) - \pi_L}{\mu_m}\right) \rightarrow 1.0$

$$R(q, d) = \frac{\log(1 + E) \cdot B_{\text{cov}} \cdot B_{\text{spec}} \cdot B_{\text{coord}} \cdot B_{\text{anc}}}{B_{\text{len}}}$$

Evidence (E):

$$E = \sum_{t \in M} qtf(t, q) \underbrace{\alpha_{qtf}}_{w(t) \text{ (term weight)}} \cdot \overbrace{\text{IDF}(t) \cdot \left(\frac{\text{IDF}(t)}{\text{IDF}(t) + 1}\right)^{\gamma_e} \cdot \left(\frac{\text{IDF}(t)}{\text{IDF}(t) + \gamma}\right)^{\beta_{tf}}}^{\text{document's coverage on the query}} \cdot \log\left(1 + \frac{tf(t, d)}{1.0}\right)$$

Coverage Multiplier ( $B_{\text{cov}}$ ):

$$B_{\text{cov}} = 1 + \gamma_{\text{cov}} \cdot \frac{\sum_{t \in M} w(t)}{\sum_{t \in q} w(t)}$$

$\xrightarrow{\text{query term weight sum}}$

Specificity Multiplier ( $B_{\text{spec}}$ ):

$$B_{\text{spec}} = 1 + \beta_{\text{spec}} \cdot \frac{1}{\sum_{t \in q} w(t)} \sum_{t \in M: \text{PMI}(t, d) > 0} w(t) \cdot \min(\text{PMI}(t, d), C_{\text{spec}})$$

3.0

$$\text{PMI}(t, d) = \log\left(\frac{tf(t, d) \cdot N}{|d|_{\text{eff}} \cdot df(t)}\right)$$

$|d|_{\text{eff}} = \max(|d|, \lfloor \text{floor} \rfloor)$   
document's focus on matched query terms beyond the corpus (in prior)

Coordination Multiplier ( $B_{\text{coord}}$ ):

$$B_{\text{coord}} = 1 + \gamma_{\text{coord}} \cdot \frac{\gamma_{\text{coord}}}{\gamma_{\text{coord}} + \log(1 + \sum_{t \in q} w(t))} \cdot \frac{|M|}{|q|}$$

$\xrightarrow{2.5}$

{ punish ~~were~~ long, common term queries  
punish less for short, rare term queries

Anchor Multiplier ( $B_{\text{anc}}$ ):

$$B_{\text{anc}} = 1 + \alpha_{\text{anc}} \cdot \log(1 + A) \quad A = \max_{t \in M: \text{IDF}(t) > \pi} \frac{\text{IDF}(t) - \pi}{\text{IDF}(t)}$$

$\pi = 4.2$

{ award the single best rare term hit

Length damper ( $B_{\text{len}}$ ):

$$B_{\text{len}} = 1 + \alpha_{\text{len}} \cdot \log\left(1 + \frac{|d| + 1}{\text{avgdl} + 1}\right) \quad \alpha_{\text{len}} = 0.15$$

{ similar to BM25, but with log, it is gentler