

# Query Likelihood Retrieval Functions

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

$$p(w | C) = \frac{c(w, C)}{\sum_{w' \in V} c(w', C)}$$

With Jelinek-Mercer (JM):

$$S_{JM}(q, d) = \log p(q | d) = \sum_{\substack{w \in d \\ w \in q}} \log \left[ 1 + \frac{1 - \lambda}{\lambda} \frac{c(w, d)}{|d| p(w | C)} \right]$$

With Dirichlet Prior (DIR):

$$S_{DIR}(q, d) = \log p(q | d) = \sum_{\substack{w \in d \\ w \in q}} \log \left[ 1 + \frac{c(w, d)}{\mu p(w | C)} \right] + n \log \frac{\mu}{|d| + \mu}$$

**What assumptions have we made in order to derive these functions?  
Do they capture the same retrieval heuristics (TF-IDF, Length Norm)  
as a vector space retrieval function?**