

Rewriting the Ranking Function with Smoothing

$$\log p(q | d) = \sum_{w \in V} c(w, q) \log p(w | d)$$

$$= \sum_{w \in V, c(w, d) > 0} c(w, q) \log p_{\text{Seen}}(w | d) + \sum_{w \in V, c(w, d) = 0} c(w, q) \log \alpha_d p(w | C)$$

Query words **matched** in d Query words **not matched** in d

$$\sum_{w \in V} c(w, q) \log \alpha_d p(w | C) + \sum_{w \in V, c(w, d) > 0} c(w, q) \log \alpha_d p(w | C)$$

All query words Query words **matched** in d

$$= \sum_{w \in V, c(w, d) > 0} c(w, q) \log \frac{p_{\text{Seen}}(w | d)}{\alpha_d p(w | C)} + |q| \log \alpha_d + \sum_{w \in V} c(w, q) \log p(w | C)$$