Rewriting the Ranking Function with Smoothing

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

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

Query words **matched** in d

Query words **not matched** in d

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

All query words

Query words **matched** in d

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