$$\text{sim}(d_j, q) = \frac{P(R|d_j)}{P(\bar{R}|d_j)}$$

Bayes: $P(A|B) = P(B|A)\,\frac{P(A)}{P(B)}$

$$= \frac{P(d_j|R)\,P(R)}{P(d_j)} \quad \frac{P(d_j)}{P(d_j|\bar{R})\,P(\bar{R})}$$

$$= \frac{P(d_j|R)}{P(d_j|\bar{R})} \quad \frac{P(R)}{P(\bar{R})} \quad \cancel{\frac{P(d_j)}{P(d_j)}} = 1$$

c

$$\sim \frac{P(d_j|R)}{P(d_j|\bar{R})}$$

Weights $w_{ij} = 1$ k_i occurs in d_j

$= 0$ k_i does not occur in d_j

words

$$\sim \prod_{w_{ij}=1} \frac{P(k_i|R)}{P(k_i|\bar{R})} \prod_{w_{ij}=0} \frac{P(\bar{k}_i|R)}{P(\bar{k}_i|\bar{R})}$$

$$\sim \prod_{w_{ij}=1} \frac{P(k_i|R)}{P(k_i|\bar{R})} \prod_{w_{ij}=0} \frac{1-P(k_i|R)}{1-P(k_i|\bar{R})}$$

Linköpings universitet

weights query $\quad w_{iq} = 1$ if k_i in query
$\quad\quad\quad\quad\quad\quad\quad\quad = 0$ if k_i not in query

$$\sim \prod_{\substack{w_{ij}=1 \\ w_{iq}=1}} \frac{P(k_i|R)}{P(k_i|\bar{R})} \prod_{\substack{w_{ij}=1 \\ w_{iq}=0}} \frac{P(k_i|R)}{P(k_i|\bar{R})}$$

$$\prod_{\substack{w_{ij}=0 \\ w_{iq}=1}} \frac{1-P(k_i|R)}{1-P(k_i|\bar{R})} \prod_{\substack{w_{ij}=0 \\ w_{iq}=0}} \frac{1-P(k_i|R)}{1-P(k_i|\bar{R})}$$

$= 1$

Assumption: terms not occurring in query are equally likely to occur in relevant documents as in non-relevant documents

$$\Rightarrow w_{iq} = 0 \Rightarrow P(k_i|R) = P(k_i|\bar{R})$$

$= 1$

$$\sim \prod_{\substack{w_{ij}=1 \\ w_{iq}=1}} \frac{P(k_i|R)}{P(k_i|\bar{R})} \frac{\prod_{\substack{w_{ij}=1 \\ w_{iq}=1}} \frac{1-P(k_i|R)}{1-P(k_i|\bar{R})}}{\prod_{\substack{w_{ij}=1 \\ w_{iq}=1}} \frac{1-P(k_i|R)}{1-P(k_i|\bar{R})}} \prod_{\substack{w_{ij}=0 \\ w_{iq}=1}} \frac{1-P(k_i|R)}{1-P(k_i|\bar{R})}$$

$$\sim \prod_{\substack{w_{ij}=1 \\ w_{iq}=1}} \frac{P(k_i|R)\,(1-P(k_i|\bar{R}))}{P(k_i|\bar{R})\,(1-P(k_i|R))}$$

$$\prod_{w_{iq}=1} \frac{1-P(k_i|R)}{1-P(k_i|\bar{R})}$$

ct for particular query

Linköpings universitet

$$\sim \prod_{\substack{w_{ij}=1 \\ w_{iq}=1}} \frac{P(k_i|R)\,(1-P(k_i|\bar{R}))}{P(k_i|\bar{R})\,(1-P(k_i|R))}$$

logarithm

$$\sim \sum_{\substack{w_{ij}=1 \\ w_{iq}=1}} \left(\log \frac{P(k_i|R)}{1-P(k_i|R)} + \log \frac{1-P(k_i|\bar{R})}{P(k_i|\bar{R})} \right)$$

$$= \sum_i w_{ij}\, w_{iq} \left(\log \frac{P(k_i|R)}{1-P(k_i|R)} + \log \frac{1-P(k_i|\bar{R})}{P(k_i|\bar{R})} \right)$$