CS598: CXZ	**Fall 2016**
Problem Set 3	
chen zhang, NetId: czhang49	*Handed In: September 24, 2016*

1 Classic Probabilistic Retrieval Model

1.1 (a)

For document generation, using multinomial model, the score can be written as:

$$\begin{aligned} score(Q,D) &= \frac{P(D|Q,R=1)}{P(D|Q,R=0)} \\ &= \frac{\prod_{j=1}^{|V|} P(w_j|Q,R=1)^{c(w_j|Q,R=1)}}{\prod_{j=1}^{|V|} P(w_j|Q,R=0)^{c(w_j|Q,R=0)}} \\ &= -\sum_{x\in\Omega} P(x)\log_2 P(x) \end{aligned}$$

And then we have

$$\begin{aligned} score(Q,D) &\propto \log \frac{P(D|Q,R=1)}{P(D|Q,R=0)} \\ &\propto \log \frac{\prod_{j=1}^{|V|} P(w_j|Q,R=1)^{c(w_j|Q,R=1)}}{\prod_{j=1}^{|V|} P(w_j|Q,R=0)^{c(w_j|Q,R=0)}} \end{aligned}$$

Since the occurrence of a word in the document is independent of the query, we have

$$c(w_j|Q,R=0) = c(w_j|Q,R=1) = c(w_j,D)$$

Then we have

$$\begin{aligned} score(Q,D) &\propto c(w_j,D)\log \frac{\prod_{j=1}^{|V|} P(w_j|Q,R=1)}{\prod_{j=1}^{|V|} P(w_j|Q,R=0)} \\ &\propto c(w_j,D)\log \prod_{j=1}^{|V|} \frac{P(w_j|Q,R=1)}{P(w_j|Q,R=0)} \\ &\propto \sum_{w\in V} c(w,D)\log \frac{P(w|Q,R=1)}{P(w|Q,R=0)} \end{aligned}$$

There's too many equations for this HW, I will write the answers very clearly starting from 1.(b). I apologize for that.