

The second step is performed at query time.

If  $q$  is a query,  $q'$  is essentially the context  
in which a search query is made

Suppose we search "bicycle" by highlighting it in  
a document. And it's surrounded by the following text

"The invention of the bicycle revolutionized personal  
transport. Early bicycles had large wheels and pedals

$q' = \text{bicycle, invention, revolutionized, personal, transport}$   
early, large, wheels, pedals

Ordinarily  $q = q'$  since it's not done in context.

$$P(C_j | q') = \frac{P(C_j) \cdot P(q' | C_j)}{P(q')} \propto P(C_j) \cdot P(q'_j | C_j)$$

$C_j$  is topic. This

is the equation used to see if a query  $q'$  belongs  
to topic  $C_j$ .  $P(C_j)$  prob of topic  $P(q'_j | C_j)$  is  
probability of term  $q'_j$  given topic  $C_j$

Imagine a query context  $q'$  is "bicycle invention wheel"

$$P(C_j | q') \propto P(C_j) \cdot P(\text{"bicycle"} | C_j) \cdot P(\text{"invention"} | C_j) \\ \cdot P(\text{"Wheels"} | C_j)$$

$$S_{qd} = \sum_j P(C_j | q') \cdot \text{rank}_{jd}$$

$\uparrow$  query sensitive score for document d given q

$\text{rank}_{jd}$  The importance score of document d  
for topic  $C_j$