

Feedback — Week 1 Practice Quiz

Help Center

Thank you. Your submission for this quiz was received.

You submitted this quiz on **Sat 21 Mar 2015 3:29 AM PDT**. You got a score of **4.00** out of **4.00**.

Question 1

Consider the instantiation of the vector space model where documents and queries are represented as **term frequency vectors**. Assume we have the following query and two documents:

Q = “future of online education”
D1 = “Coursera is shaping the future of online education; online education is affordable.”
D2 = “In the future, online education will dominate.”

Let $V(X) = [c_1 \ c_2 \ c_3 \ c_4]$ represent a part of the term frequency vector for document or query X, where c_1 , c_2 , c_3 , and c_4 are the term weights corresponding to “future”, “of”, “online”, and “education”, respectively. Which of the following is true:

Your Answer	Score	Explanation
<input checked="" type="radio"/> $V(Q) = [1 \ 1 \ 1 \ 1]$ $V(D1) = [1 \ 1 \ 2 \ 2]$ $V(D2) = [1 \ 0 \ 1 \ 1]$	✓ 1.00	
<input type="radio"/> $V(Q) = [1 \ 1 \ 1 \ 1]$ $V(D1) = [1 \ 1 \ 1 \ 1]$ $V(D2) = [1 \ 0 \ 1 \ 1]$		
<input type="radio"/> $V(Q) = [1 \ 1 \ 1 \ 1]$ $V(D1) = [1 \ 1 \ 2 \ 2]$ $V(D2) = [1 \ 1 \ 1 \ 1]$		
Total	1.00 / 1.00	

Question 2

Consider the same scenario as in question (1) with the dot product as the similarity measure.

Which of the following is the true:

Your Answer	Score	Explanation
<input type="radio"/> $\text{Sim}(Q,D1) = 6 \quad \text{Sim}(Q,D2) = 4$		
<input checked="" type="radio"/> $\text{Sim}(Q,D1) = 6 \quad \text{Sim}(Q,D2) = 3$	✓ 1.00	
<input type="radio"/> $\text{Sim}(Q,D1) = 4 \quad \text{Sim}(Q,D2) = 3$		
Total	1.00 / 1.00	

Question 3

Assume we have two documents with the same raw TF for all the query words (i.e. the query words appear with the same frequency in both documents). Then, using the Okapi/BM25 retrieval function, the longer document will have a lower score.

Your Answer	Score	Explanation
<input checked="" type="radio"/> True	✓ 1.00	
<input type="radio"/> False		
Total	1.00 / 1.00	

Question 4

If we remove the document length normalization term from the Okapi/BM25 retrieval function, and have two documents with the same raw TF for all the query words, then the longer document will have a higher score.

Your Answer	Score	Explanation
<input type="radio"/> True		
<input checked="" type="radio"/> False	✓ 1.00	
Total	1.00 / 1.00	

