Close

Week 1 Quiz

10 questions

1  
point

1.

The sentence “A man saw a boy with a telescope” is syntactically ambiguous and has two distinct syntactic structures.



False



True

1  
point

2.

Which of the following is false?



Querying and browsing are both examples of the text pull mode.



Search engines rely on the text push mode.



Recommender systems are based on the text push mode.



Browsing is suitable when the user doesn't know what keywords to use.

1  
point

3.

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

Q = “healthy diet plans”

D1 = “healthy plans for weight loss. Check out other healthy plans”

D2 = “the presidential candidate plans to change the educational system.”

Let V(X) = [b1 b2 b3] represent a part of the bit vector for document or query X, where b1, b2, and b3 are the bits corresponding to “healthy,” “diet,” and “plans,” respectively.

Which of the following is true?



V(Q) = [1 1 1]  V(D1) = [1 1 1]  V(D2) = [0 0 1]



V(Q) = [1 1 1]  V(D1) = [1 1 1]  V(D2) = [0 0 0]



V(Q) = [1 1 1]  V(D1) = [2 0 2]  V(D2) = [0 0 1]



V(Q) = [1 1 1]  V(D1) = [1 0 1]  V(D2) = [0 0 1]

1  
point

4.

Consider the same scenario as in Question 3, with dot product as the similarity measure. Which of the following is true?



Sim(Q,D1) = 2  Sim(Q,D2) = 1



Sim(Q,D1) = 3  Sim(Q,D2) = 0



Sim(Q,D1) = 3  Sim(Q,D2) = 1



Sim(Q,D1) = 4  Sim(Q,D2) = 1

1  
point

5.

In the "simplest" VSM instantiation, if instead of using 0-1 bit vectors but we use the word count instead, when we concatenate each document by itself, will the ranking list still remain the same?



True



False

1  
point

6.

In Text Retrieval problem for N distinct documents, select statements below that are correct?



Document selection is preferred as there is no need to determine document absolute relevance



If use document selection, the number of outcome is 2*N*



If use document ranking, the number of outcome is *N*!



The numbers of outcome for document ranking and selection are the same

1  
point

7.

Suppose we compute the term vector for a baseball sports news article in a collection of general news articles using **TF weighting only**. Which of the following words do you expect to have the highest weight?



the



baseball



computer

1  
point

8.

Assume the same scenario as in Question 7, but with **TF-IDF weighting**. Which of the following words do you expect to have the highest weight in this case?



the



baseball

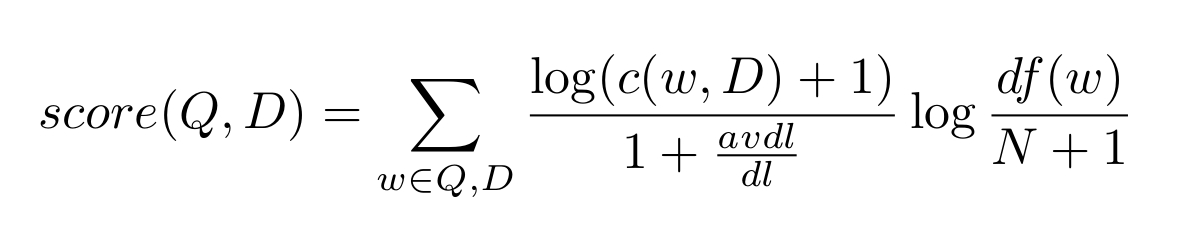


computer

1  
point

9.

Consider the following retrieval formula:



Where c(w, D) is the count of word w in document D,

dl is the document length,

avdl is the average document length of the collection,

N is the total number of documents in the collection,

and df (w) is the number of documents containing word w.

In view of TF, IDF weighting, and document length normalization, which part is missing or does not work appropriately?



TF



Document length normalization



IDF

1  
point

10.

When using the Okapi/BM25 retrieval function on a corpus where each document has exactly the same length, removing the document length normalization term from the retrieval function will change the ranking of documents.



False



True

Submit Quiz