START OF QUIZ
Student ID: 60226669,Mohammad,Waquas

Question 1

Topic: Lecture 8
Source: Lecture 8

What is the reasonining behind substituting TF-IDF with Okapi BM25? (1)

Question 2

Topic: Lecture 5
Source: Lecture 5

The Frobenius norm looks very similar to a distance metric we've already observed. Explain which one. (1)

Question 3

Why can we be confident that a low-rank approximation of a matrix contains the most important information in a document? (1)

Question 4

Explain why the cosine similarity between a document and query vector is roughly equivalent to adding up the TF-IDF scores of each word in the document that occurs in the query.
(2)

Question 5

Imagine we performed LDA on the classes in this block. What might their [Beta / Theta] distributions look like? (2)

Question 6

Topic: Lecture 8
Source: Lecture 8

What are some assumptions that we make when we are interpolating between a document and a corpus? When should we trust the corpus more, and when should we trust the document more? (2)

Question 7

Topic: Lecture 6
Source: Lecture 6

Why do we need a "human in the loop" for topic modeling? (1)

Question 8

Topic: Lecture 7
Source: Lecture 7

What is the purpose of an inverted index? (1)

Question 9

Topic: Long
Source: Lecture 5

Imagine that we are working with a language other than English, such as Indonesian, with significant agglutinative morphology (words are inflected through the concatenation of affixes to a lemma). How do you think that this would impact our various vector space models? Which of them would be most affected, and which would be least affected? Explain. (3)

END OF QUIZ