START OF QUIZ
Student ID: 96158571,Bandaru,Sai Charan

Question 1

Topic: Lecture 7
Source: Lecture 7

What is the benefit of evaluating boolean queries using set operations instead of loops? (1)

Question 2

Topic: Lecture 7
Source: Lecture 7

Explain why boolean filtering is usually insufficient for retrieval, and why we normally need some way of scoring the documents. (2)

Question 3

Topic: Lecture 8
Source: Lecture 8

Why do we not simply take the probability of a word given its document (maybe with smoothing added in)? (1)

Question 4

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 5

Topic: Lecture 5
Source: Lecture 5

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

Question 6

Topic: Lecture 6
Source: Lecture 6

In class, we talked about bookstores and streaming algorithms classifying books / movies. How can we tell that they don't use a topic modeling algorithm (or, if you think they do, what would be some clues)? (1)

Question 7

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 8

Topic: Lecture 6
Source: Lecture 6

In some ways, we could consider Theta distributions themselves to be an embedding of a topic. Explain, and explain how we might be able to leverage that. (2)

Question 9

Topic: Coding
Source: Coding

Write a short function that confirms that the sum of n rank-1 matrices is identical to the product of an nxk matrix and a kxn matrix. (3)

END OF QUIZ