START OF QUIZ
Student ID: 34157719,Philip,Reshmi

Question 1

Topic: Lecture 5
Source: Lecture 5

We often weight our matrices using something like PMI or TF-IDF. Do you think it would make sense to do this after applying SVD? Why or why not? (2)

Question 2

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 3

Topic: Lecture 8
Source: Lecture 8

What do we mean by interpolation? (1)

Question 4

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

Question 5

Explain the logic behind the IDF part of TF-IDF (ie, why does it give higher weights to more "interesting" words?). (1)

Question 6

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

Question 7

Why do we generally care more about precision than recall in IR? (1)

Question 8

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 9

Topic: Long
Source: Lecture 7

Imagine that we have 2 information retrieval systems, and we are evaluating on the same test set, which has 10 relevant documents. The first system returns them in positions [1, 5, 7, 15, 25, 50, 60, 70, 71, 90]. The second returns the documents at positions [2, 3, 6, 8, 10, 62, 80, 83, 91, 95]. Make an argument for each system being better, and provide support for both. Explain which system you would rather use, and why. If there are any other considerations, list them. (3)

END OF QUIZ