START OF QUIZ
Student ID:
97168009,Shankar,Ashwin

Question 1

Topic: Lecture 7
Source: Lecture 7

What is the benefit (in terms of efficiency) of placing the most discriminative search terms first in a boolean search? (1)

Question 2

Topic: Lecture 8
Source: Lecture 8

Why don't we use a higher-order language model to perform IR? (1)

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 6
Source: Lecture 6

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

Question 5

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 6

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 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

Why don't we just use k-means to cluster document-vectors (sparse or dense)? (1)

Question 9

Topic: Coding
Source: Coding

Write a function that returns the most likely n documents given a term-document matrix, a smoothing parameter, and a query. (3)

END OF QUIZ