START OF QUIZ
Student ID:
86330909,Xia,Liangchen

Question 1

Topic: Lecture 5
Source: Lecture 5

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

Question 2

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 3

Topic: Lecture 7
Source: Lecture 7

What is the benefit of evaluating boolean queries using set operations instead of loops? (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 6
Source: Lecture 6

Why can't we just run an HMM over documents to discover the latent states like we do for POS-tagging? (1)

Question 6

Topic: Lecture 7
Source: Lecture 7

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

Question 7

Topic: Lecture 8
Source: Lecture 8

What do we mean by interpolation? (1)

Question 8

Topic: Lecture 5
Source: Lecture 5

Why can we represent a rank-m matrix as the sum of m rank-1 matrices *or* the product of an n x m matrix and an m x n matrix (ie, what is matrix multiplication doing that we can take advantage of?)? Explain. (2)

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