START OF QUIZ
Student ID:
74874876,Wang,Chenxin

Question 1

Topic: Lecture 5
Source: Lecture 5

What advantages do sparse vectors have over dense ones? (1)

Question 2

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 3

Topic: Lecture 8
Source: Lecture 8

In class, I mentioned that high k value for BM25 TF weigthing rewards documents with many, many instances of a term in them. Explain why that's the case. (2)

Question 4

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 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 8
Source: Lecture 8

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

Question 7

Topic: Lecture 6
Source: Lecture 6

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

Question 8

Topic: Lecture 7
Source: Lecture 7

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

Question 9

Topic: Long
Source: Lecture 8

In class, we considered two different types of information retrieval systems - one that uses Boolean terms to find matches, and one that uses a language model to allow for "natural language" queries. Can you think of a way that we might be able to leverage the strengths of both, while minimizing the disadvantages? Briefly explain how that might work. (2)

END OF QUIZ