START OF QUIZ
Student ID:
84574284,Cheng,Yushun

Question 1

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

Question 2

Topic: Lecture 8
Source: Lecture 8

What do we mean by interpolation? (1)

Question 3

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 4

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

Question 5

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 6

Topic: Lecture 7
Source: Lecture 7

Explain why the cosine similarity between a document and query vector is roughly equivalent to adding up the TF-IDF scores of each word in the document that occurs in the query. (2)

Question 7

Topic: Lecture 7
Source: Lecture 7

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

Question 8

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 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