START OF QUIZ
Student ID:
19668508,Li,Julian

Question 1

Topic: Lecture 8
Source: Lecture 8

In class, I mentioned that TF-IDF unfairly punishes words that appear in a lot, but not all, of the documents in our corpus. Explain how Okapi BM25 attempts to fix this. (1)

Question 2

Topic: Lecture 6
Source: Lecture 6

Why do we need a "human in the loop" for topic modeling? (1)

Question 3

Topic: Lecture 8
Source: Lecture 8

What do we mean by interpolation? (1)

Question 4

Topic: Lecture 7
Source: Lecture 7

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

Question 5

Topic: Lecture 5
Source: Lecture 5

What impact do sparse matrices have on similarity metrics like cosine similarity? (1)

Question 6

Topic: Lecture 7
Source: Lecture 7

When doing information retrieval, bag-of-words (and even just indicator functions) typically work very well. Explain why context is less important if we have a well-designed query. You may also want to explain your assumptions about a "well-designed" query. (2)

Question 7

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 8

Topic: Lecture 6
Source: Lecture 6

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

Question 9

Topic: Long
Source: Lecture 8

In class, I mentioned that we might want to provide a weighted ranking of documents in an IR system (for example, we might have a list of relevant documents already sorted for relevancy, and we want our IR system to not only return those documents high in the returned document list, but in the same order). How might we modify the MAP metric to reward the ordering of the documents, as well? Explain (use an example, if you have to). (3)

END OF QUIZ