START OF QUIZ
Student ID:
37157856,Wang,Lusha

Question 1

Topic: Lecture 8
Source: Lecture 8

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

Question 2

Topic: Lecture 6
Source: Lecture 6

In class, we saw a few topics that we were unable to identify. What could be a cause for such pointless topics (ie, how might we ensure that our topics are better? (2 reasons). (1)

Question 3

Topic: Lecture 7
Source: Lecture 7

From a processing perspective, what is one benefit structured data has over unstructured data, and vice versa. (1)

Question 4

Topic: Lecture 8
Source: Lecture 8

What do we mean by interpolation? (1)

Question 5

Topic: Lecture 5
Source: Lecture 5

What advantages do sparse vectors have over dense ones, and vice versa? (1)

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 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 8

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 9

Topic: Long
Source: Lecture 5

Imagine that we are working with a language other than English, such as Indonesian, with significant agglutinative morphology (words are inflected through the concatenation of affixes to a lemma). How do you think that this would impact our various vector space models? Which of them would be most affected, and which would be least affected? How might we go about adapting our model to solve these problems? Explain. (3)

END OF QUIZ