START OF QUIZ
Student ID:
33399049,Cross,Ziggy

Question 1

Topic: Lecture 6
Source: Lecture 6

We've seen that seed lexicons can help with semisupervised learning. Explain how we might be able to use LDA to derive a translation seed lexicon. (2)

Question 2

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 3

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 4

Topic: Lecture 8
Source: Lecture 8

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

Question 5

Topic: Lecture 7
Source: Lecture 7

Define P@R. (1)

Question 6

Topic: Lecture 5
Source: Lecture 5

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

Question 7

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 8

Topic: Lecture 8
Source: Lecture 8

What do we mean by interpolation? (1)

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