START OF QUIZ
Student ID:
77026334,Qiassi,Behrooz

Question 1

Topic: Lecture 7
Source: Lecture 7

Explain the underlying assumption of the TextTiling algorithm. (1)

Question 2

Topic: Lecture 8
Source: Lecture 8

Describe a Discourse Unit. (1)

Question 3

Topic: Lecture 6
Source: Lecture 6

When running a window-based approach to vector embeddings (such as CBOW or skip-gram), when would it make sense to keep stopwords, and when would it make sense to remove them? (1)

Question 4

We took a look at how vectors can be added / subtracted in vector space. Why does this work? (hint: think back to the general properties of word embeddings that we've wanted from the very start) (1)

Question 5

The TextTiling algorithm we looked at just looked at raw word overlap (possibly with stop-word removal). Describe a way that we could improve the algorithm to maximize coverage. (2)

Question 6

Topic: Lecture 5
Source: Lecture 5

Which is likely to have the lowest PMI? A rare word and a frequent word that appear together frequently, or two frequent words that appear together frequently? (1)

Question 7

In class, we talked about how a "typical" dimensionality for embeddings is in the range of 100-500. What might be some consequences if we estimated too low or too high? (2)

Question 8

Topic: Lecture 8
Source: Lecture 8

Do you think we could use word embeddings for coreference resolution? What kind of assumptions would we be making, and why do you think it might still be a very difficult task? (2)

Question 9

Write a short function that tries to find a good value for k in truncated SVD. You'll essentially be writing your own version of the evaluate word analogies function. For each of our the analogies, you'll need to do the vector math we were doing in the capital city determination, and return the 1-closest vector - if it's what we're looking for, it's correct. If not, it's wrong. (3)

END OF QUIZ