START OF QUIZ
Student ID: 75287953,Apparaju,Ananya

Question 1

Topic: Lecture 5
Source: Lecture 5

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 2

Topic: Lecture 8
Source: Lecture 8

Why is it easy to create negative examples for lexical coherence tests? (1)

Question 3

Topic: Lecture 7
Source: Lecture 7

How is the TextTiling algorithm similar to the Lesk algorithm? How is it different? (2)

Question 4

Topic: Lecture 8
Source: Lecture 8

What is an anaphor? (1)

Question 5

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 6

Topic: Lecture 7
Source: Lecture 7

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

Question 7

Topic: Lecture 6
Source: Lecture 6

Think back to week 1 of this block when we were doing word sense disambiguation. Do you think there would be benefits to disambiguating all words before running word2vec? Explain. (2)

Question 8

Topic: Lecture 5
Source: Lecture 5

Generally speaking, why are we not interested in negative PMI? (1)

Question 9

Topic: Coding
Source: Coding

Imagine we were trying to find a word that is the best prototype of its synonyms. Write a short function that grabs the lemmas of each synset in wordnet, and calculates which lemma is the best prototype (ie, which lemma is the closest to the centroid of the synset) by using the word embeddings. Ignore words that do not have embeddings in gensim. (3)

END OF QUIZ