START OF QUIZ
Student ID:
74035403,Zheng,Meiyu

Question 1

Topic: Lecture 7
Source: Lecture 7

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 2

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 3

Topic: Lecture 6
Source: Lecture 6

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 4

Topic: Lecture 8
Source: Lecture 8

Describe the recency criterion for anaphor resolution. Why can't we just backtrack from the current word (at least in English)? (2)

Question 5

We've seen co-occurrence matrices weighted by TF-IDF- would it make sense to weight them by PMI? Briefly explain. (1)

Question 6

Topic: Lecture 5
Source: Lecture 5

What is PMI measuring? That is, what does it mean for two words to have high PMI? (1)

Question 7

Topic: Lecture 8
Source: Lecture 8

What is an anaphor? (1)

Question 8

Topic: Lecture 7
Source: Lecture 7

Explain the underlying assumption of the TextTiling algorithm. (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