START OF QUIZ
Student ID:
26566455,Lai,Minsi

Question 1

Topic: Lecture 7
Source: Lecture 7

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

Question 2

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 3

Topic: Lecture 5
Source: Lecture 5

What is the primary assumption of the vector space model for semantics, regardless of how it's implemented? (1)

Question 4

Topic: Lecture 6
Source: Lecture 6

Explain why extrinsic evaluation can be a much more desirable method of evaluating the quality of word vectors than intrinsic evaluation (we didn't have this in the slides, but remember that intrisic evaluation is something like the analogy task, that tries to measure the quality of the vectors directly). (2)

Question 5

Topic: Lecture 7
Source: Lecture 7

Why are we interested in backward-facing centers (Cb)? Why not just consider the entities in the current sentence? (1)

Question 6

Topic: Lecture 8
Source: Lecture 8

Describe a Discourse Unit. (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 or disadvantages to disambiguating all words before running word2vec? Explain. (2)

Question 8

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 9

Topic: Long
Source: Lecture 6

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