

START OF QUIZ

Student ID:

34447581,Ong,Claudia

Question 1

Topic: Lecture 5

Source: Lecture 5

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

Question 2

Topic: Lecture 6

Source: Lecture 6

What are two significant shortcoming of the Word2Vec model? (1)

Question 3

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 4

Topic: Lecture 7

Source: Lecture 7

Why do you think that pronouns must be high salience items from previous sentences? (1)

Question 5

Topic: Lecture 8

Source: Lecture 8

Describe a Discourse Unit. (1)

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

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 7

Source: Lecture 7

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