

START OF QUIZ

Student ID:

37289428, Yun, Michelle

Question 1

Topic: Lecture 8

Source: Lecture 8

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

Question 2

Topic: Lecture 7

Source: Lecture 7

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

Question 3

Topic: Lecture 6

Source: Lecture 6

Why do we say that the analogy task is an "intrinsic" evaluation of our word embeddings? (1)

Question 4

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 5

Topic: Lecture 6

Source: Lecture 6

What is the purpose of negative sampling in a Word2Vec model? (1)

Question 6

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 7

Topic: Lecture 5

Source: Lecture 5

When we were calculating PMI of a symmetric matrix, why is it not a case of double counting the word in our document? ie., why do the counts of (attorney, fun) and (fun, attorney) not count as two counts each of attorney and fun (such as when we are calculating the total sum of the matrix? (2)

Question 8

Topic: Lecture 8

Source: Lecture 8

What tools are required to build an entity grid? (not structures - matrices, etc. are interesting, but I'm asking what kind of NLP tools are necessary to fill the grid - there are at least 2. (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