START OF QUIZ
Student ID:
32029357,Oh,Jacob

Question 1

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 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 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 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 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

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

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