START OF QUIZ
Student ID:
43887546,Kumar,Abhi

Question 1

Topic: Lecture 6
Source: Lecture 6

When running a window-based approach to vector embeddings (such as CBOW or skip-gram), when would it make sense to keep stopwords, and when would it make sense to remove them? (1)

Question 2

Topic: Lecture 7
Source: Lecture 7

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

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

Describe a Discourse Unit. (1)

Question 5

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 6

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 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

What is an anaphor? (1)

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