

**START OF QUIZ**

**Student ID:**

**30425177,He,Hao Chen**

## 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

We took a look at 2 different ways of implementing the TextTiling algorithm - one with vector overlap, and one with BERT. Can you think of how we might modify the algorithm further to strengthen up its weaknesses? (No is not a valid answer.) (2)

### Question 3

Topic: Lecture 6

Source: Lecture 6

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

## Question 4

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 5

Topic: Lecture 5

Source: Lecture 5

In class, we talked about how a "typical" dimensionality for embeddings is in the range of 100-500. What might be some consequences if we estimated too low or too high? (2)

## Question 6

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 7

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 8

Topic: Lecture 6

Source: Lecture 6

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