START OF QUIZ
Student ID:
96733092,Zhou,Zhiyang

Question 1

Explain the underlying assumption of the TextTiling algorithm. (1)

Question 2

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 3

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 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 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 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 8
Source: Lecture 8

Do you think we could use word embeddings for coreference resolution? What kind of assumptions would we be making, and why do you think it might still be a very difficult task? (2)

Question 8

We took a look at how vectors can be added / subtracted in vector space. Why does this work? (hint: think back to the general properties of word embeddings that we've wanted from the very start) (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