

**START OF QUIZ**

**Student ID:**

**88683578,Mao,MingJia**

## Question 1

Topic: Lecture 7

Source: Lecture 7

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

## Question 2

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 3

Topic: Lecture 8

Source: Lecture 8

What is an anaphor? (1)

## Question 4

Topic: Lecture 5

Source: Lecture 5

Generally speaking, why are we not interested in negative PMI? (1)

## Question 5

Topic: Lecture 6

Source: Lecture 6

Explain why extrinsic evaluation can be a much more desirable method of evaluating the quality of word vectors than intrinsic evaluation (we didn't have this in the slides, (2)

## Question 6

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 disadvantages to disambiguating all words before running word2vec? Explain. (2)

## Question 7

Topic: Lecture 5

Source: Lecture 5

Which is likely to have the lowest PMI? A rare word and a frequent word that appear together frequently, or two frequent words that appear together frequently? (1)



## 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

Identify the shifts in the following discourse (show your work): Jonathan Harker was a solicitor from England. He was sent to Transylvania to meet with the mysterious Count Dracula. Dracula wanted to buy property in London. That's where all the wealthiest nobles lived. Dracula had other plans, too, but Harker didn't know that. (3)

**END OF QUIZ**