START OF QUIZ
Student ID:
30821250,Huang,Chloe

Question 1

Topic: Lecture 4
Source: Lecture 4

Why does type-to-token ratio decrease as the size of the corpus increases? What does this suggest about long documents? (1)

Question 2

Topic: Lecture 3
Source: Lecture 3

Although lexicons are often good starting points, they are often less capable than ML methods. What are some reasons (at least 2) that lexicons are insufficient for state-of-the-art training. Briefly explain. (2)

Question 3

Topic: Lecture 2
Source: Lecture 2

Would a language with with lots of inflection have a higher or lower TTR than one with little inflection? Briefly explain. (1)

Question 4

Topic: Lecture 3
Source: Lecture 3

Imagine that we have a parallel corpus (ie, a corpus containing sentences in two languages), and we want to extract a bilingual lexicon. What are some simple steps we could do to identify words that could be translations of each other? (2)

Question 5

Topic: Lecture 1
Source: Lecture 1

Vowels are often used as a proxy for syllables in words (it's not a perfect correspondence, but it's not bad). Write a function that counts the vowels in a word, without using a loop, using only the tools we went over in Lecture 1 (list comprehension counts as a loop). (2)

Question 6

Topic: Lecture 4
Source: Lecture 4

Do you think that children's (age 3-5) picture books would have a higher or lower ratio of adjectives than university literature? Briefly explain your logic. (1)

Question 7

Topic: Lecture 2
Source: Lecture 2

What are 2 benefits of providing a corpus reader with your corpus? (1)

Question 8

Topic: Lecture 1
Source: Lecture 1

When would we *not* want to lowercase text prior to training a model? Give a concrete example. (1)

Question 9

Topic: Coding
Source: Lecture 2

Imagine we have a large corpus in an unknown language. We don't have any ML tools to analyze the data. How might we determine the stopwords in our corpus? How might we test our theory of stopwords? (I'll make it easy on you - the tokens are space separated, and we have some way of separating sentences.) (3)

END OF QUIZ