

START OF QUIZ

Student ID:

80162985,Sirwani,kartik

Question 1

Topic: Lecture 4

Source: Lecture 4

Would you expect a higher or lower frequency of passive voice constructions in legal documents compared to casual conversation? Briefly explain your reasoning. (Remember that passive voice is a structure like "the tree was cut down", inverting the subject and object). (1)

Question 2

Topic: Lecture 2

Source: Lecture 2

As we increase the size of a corpus, the frequency of Hapax Legomena generally increases. Would the frequency of function words like "the" or "is" also increase? Why or why not?
(1)

Question 3

Topic: Lecture 1

Source: Lecture 1

What are two ways to check if a string is a palindrome, without reversing the string? (1)

Question 4

Topic: Lecture 4

Source: Lecture 4

Given a list of tuples where the first element is a string and the second is an integer, write a short piece of code to sort the list in descending order based on the second element. Briefly explain your approach. (1)

Question 5

Topic: Lecture 2

Source: Lecture 2

Why is it important to understand the intended audience and time period of a corpus when conducting linguistic analysis? (1)

Question 6

Topic: Lecture 3

Source: Lecture 3

Imagine you have a large text corpus in English and Spanish and want to automatically align sentences for machine translation. What are some straightforward methods you could use to identify sentence pairs that are likely translations of each other? (2)

Question 7

Topic: Lecture 3

Source: Lecture 3

Lexicons are useful for initial text analysis but often lack the adaptability needed for advanced NLP tasks. Why is this the case? Provide at least 2 reasons with brief explanations.
(2)

Question 8

Topic: Lecture 1

Source: Lecture 1

Write a function that capitalizes the first letter of each word in a string, without using the `.title()` method or any external libraries. What are some assumptions that you are making? (2)

Question 9

Topic: Long

Source: Lecture 4

Suppose you have a large corpus of mixed language text, and you are tasked with detecting which language each sentence is written in. What techniques could you apply to achieve this, assuming you cannot use a pre-trained language detection model? Please provide a detailed explanation of your reasoning. (3)

END OF QUIZ