

START OF QUIZ

Student ID:

**70152863,Khan,Muhammad
Mujtaba**

Question 1

Topic: Lecture 4

Source: Lecture 4

Why does the lexical diversity (type-to-token ratio) typically increase when analyzing smaller sub-corpora rather than larger ones? What does this suggest about the content of smaller texts? (1)

Question 2

Topic: Lecture 3

Source: Lecture 3

When we nest deep structures in dictionaries, we lose their $O(1)$ benefits. Can you think of a better way to represent complex data sets? (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 3

Source: Lecture 3

Why do we not care about the extra space required to create a reverse index? (2 reasons) (1)

Question 5

Topic: Lecture 2

Source: Lecture 2

How does Zipf's law help explain the distribution of word frequencies in a corpus? What impacts does that have on our algorithms? (1)

Question 6

Topic: Lecture 4

Source: Lecture 4

Attributive adverbs are a type of adverb that provides "flavour" to speech verbs (example: "she said quickly"; "he spoke loudly"). They are often frowned upon in formal writing, because they can be replaced with other verbs: "blurted" or "shouted", in the example. Write a quick function that finds them in the Brown corpus, and reports how many sentences in 1000 have them. (2)

Question 7

Topic: Lecture 1

Source: Lecture 1

You are given a sentence. Write a function to count how many words in the sentence start with a vowel, without using loops or list comprehensions. (2)

Question 8

Topic: Lecture 2

Source: Lecture 2

If you were to analyze a corpus for stylistic differences, how might you determine: the formality of the language; whether it's written or spoken; its sentiment? Assume that we don't have existing ML tools or enough data to train one. (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