START OF QUIZ
Student ID:
97713317,Song,Shawn

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

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

What is the Big O time complexity of finding the elements in a set that intersect with an iterable (ie, string, list, etc)? Briefly explain. (1)

Question 5

Topic: Lecture 1
Source: Lecture 1

What method would you use to check if a string contains only numeric digits (including decimals) without using any additional libraries? (1)

Question 6

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 7

Topic: Lecture 2
Source: Lecture 2

Is it possible for a corpus of a low-resource language to follow Zipf's law? What factors might influence the degree to which the law applies in such languages? (2)

Question 8

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 9

Topic: Long
Source: Lecture 1

Write a function that validates if a string matches a phone number format, such as (123) 456-7890. What types of invalid inputs should the function check for? Are there edge cases we would be willing to accept? How would we handle those? Write 3 test cases - 2 that should pass, and one that should fail. (3)

END OF QUIZ