START OF QUIZ
Student ID:
37157856,Wang,Lusha

Question 1

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 2

What role does linguistic annotation provide for corpora, specifically for computational linguistics? (1)

Question 3

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

Question 4

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 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 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

In French, negation is often indicated by "ne ... pas" (ie, "je ne parle pas" - "I am not speaking"; "tu ne conduis pas" - "You are not driving", etc.). However, in speech, one of the two is often dropped: "je ne parle." or "tu conduis pas.". Using this information, how would you determine whether a corpus was composed of written or spoken French? You don't need to write the code, but explain the logic that you would use to come to this conclusion. (2)

Question 8

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 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