START OF QUIZ
Student ID:
55343529,Nadal,Jacob

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 1
Source: Lecture 1

When would you choose to preserve the original case of text during data processing, rather than converting everything to lowercase? (1)

Question 3

Topic: Lecture 1
Source: Lecture 1

How would you convert a string into a list of characters? (1)

Question 4

Topic: Lecture 3
Source: Lecture 3

When would we want to represent linguistic data in a list, instead of a dictionary or a set? (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 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 7

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 8

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 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