

START OF QUIZ

Student ID:

52196086,Ahluwalia,Max

Question 1

Topic: Lecture 1
Source: Lecture 1

Vowels are often used as a proxy for syllables in words (it's not a perfect correspondence, but it's not bad). Write a function that counts the vowels in a word, without using a loop, using only the tools we went over in Lecture 1 (list comprehension counts as a loop). (2)

Question 2

Topic: Lecture 3

Source: Lecture 3

Although lexicons are often good starting points, they are often less capable than ML methods. What are some reasons (at least 2) that lexicons are insufficient for state-of-the-art training. Briefly explain. (2)

Question 3

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 4

Topic: Lecture 4

Source: Lecture 4

Why does type-to-token ratio decrease as the size of the corpus increases? What does this suggest about long documents? (1)

Question 5

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 6

Topic: Lecture 1

Source: Lecture 1

Why is `strip()` such a useful function? (1)

Question 7

Topic: Lecture 2

Source: Lecture 2

How does Zipf's law relate to Hapax Legomena? (1)

Question 8

Topic: Lecture 2

Source: Lecture 2

Would a language with with lots of inflection have a higher or lower TTR than one with little inflection? Briefly explain. (1)

Question 9

Topic: Coding

Source: Coding

Imagine that we have an encrypted data set in a language we don't know, but it is written in the Latin script (ie, the script of English, French, etc.). What are some tests that we could run to try to determine the original language? Please list any assumptions you make. Assume that machine learning is not an option. (3)

END OF QUIZ