START OF QUIZ
Student ID:
83261909,Dang,Trang

Question 1

Topic: Lecture 1
Source: Lecture 1

How would you reverse a string and keep the result? (1)

Question 2

Topic: Lecture 2
Source: Lecture 2

Do you think it's possible for a language not to follow a Zipfian curve? What consequences might that have on communication (if, let's say, if the curve were linear)? (2)

Question 3

Topic: Lecture 2
Source: Lecture 2

If we have a new corpus, how might we automatically determine (without ML): A. The language it's written in. B. Whether it is annotated C. If it is multilingual D. genre? Briefly explain your reasoning. (2)

Question 4

Topic: Lecture 1
Source: Lecture 1

When would we *not* want to lowercase text prior to training a model? Give a concrete example. (1)

Question 5

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 6

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 7

Topic: Lecture 3
Source: Lecture 3

What properties of dictionaries make them an efficient choice for nesting complex lexicons. (1)

Question 8

Topic: Lecture 4
Source: Lecture 4

How would we sort a dictionary alphabetically by the reverse of its keys (assuming the keys are strings)? Write a short piece of code, and briefly explain your logic. (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