START OF QUIZ
Student ID: 56633415,Cady,Amandisa

Question 1

Topic: Lecture 7
Source: Lecture 7

Can you think of any classes of words in English where the stem and the lemma will always be identical? Why is that of little interest to us? (1)

Question 2

Topic: Lecture 8
Source: Lecture 8

If you were to encounter an alien text, which encoding might you want to use to digitize it? Explain briefly. (1)

Question 3

Topic: Lecture 5
Source: Lecture 5

List one advantage that regular expressions have over string comparison, and one disadvantage to using them. (1)

Question 4

Topic: Lecture 6
Source: Lecture 6

Why is XML well-suited to representing linguistic data? (1)

Question 5

Topic: Lecture 8
Source: Lecture 8

Why should you get into the habit of using ”with open()”? Are there any downsides? (1)

Question 6

Topic: Lecture 7
Source: Lecture 7

I mentioned in class that POS tagging is often viewed as a pre-processing step for many CL tasks. What assumptions are we making (at least 3) when including it in our NLP pipeline? Do you think these are reasonable assumptions, and if they fail, is it worth the effort to solve the problem, or just ignore POS tagging? (2)

Question 7

Topic: Lecture 6
Source: Lecture 6

Consider using XML to represent a machine learning model's architecture. What XML tags might be useful for representing layers, activation functions, and connections between layers (you don't need to describe a deep-learning architecture - describe one you're familiar with)? If this doesn't seem possible, explain why not. (2)

Question 8

Topic: Lecture 5
Source: Lecture 5

Imagine you are processing a text document where dates are written in multiple formats, such as ”12-05-2024”, ”05/12/2024”, or ”12 December 2024”. How would you write a regex to capture these date formats (just the logic)? What assumptions would you make? (2)

Question 9

Topic: Long
Source: Lecture 7

Suppose you're building a text classification model for a highly inflected language like Finnish. How might you approach preprocessing tasks such as lemmatization or stemming? Would you perform these tasks before or after feature extraction, and why? Discuss how the choice of sequence may impact the quality of the features and model accuracy. Would you make the same decision for sentiment analysis? (3)

END OF QUIZ