START OF QUIZ
Student ID:
72119878,Wang,Yuanao

Academic honesty is essential to the continued functioning of the University of British Columbia as an institution of higher learning and research. All UBC students are expected to behave as honest and responsible members of an academic community. Failure to follow the appropriate policies, principles, rules, and guidelines of the University with respect to academic honesty may result in disciplinary action.

I agree that all answers provided are in my own words, and that I will not discuss the contents of this quiz with any of my fellow students until after the exam period has completed for everyone. Furthermore, any response that used generative AI tools has been rephrased into my own interpretation, and has been appropriately cited.

Signature: _____

Question 1

Topic: Lecture 5
Source: Lecture 5

There are two ways of matching a pattern against the start of a string. Describe them. (1)

Question 2

Topic: Lecture 5
Source: Lecture 5

Write a regex pattern that matches any valid email address (i.e., with basic rules like user@domain.com). What challenges might you face in accurately matching all possible email formats? (1)

Question 3

Topic: Lecture 7
Source: Lecture 7

What might the training data for a sentence segmenter look like? Do you think it would be easy or hard to train? Explain briefly. (1)

Question 4

Topic: Lecture 8
Source: Lecture 8

Why do Python programmers like working with (t/c)sv files? When are they appropriate, and what advantages do they provide over .txt files? (1)

Question 5

Topic: Lecture 6
Source: Lecture 6

What kinds of tags might be useful in the following text (describe at least two): ”But you liked Rashomon!” ”That’s not how I remember it!” (1)

Question 6

Topic: Lecture 7
Source: Lecture 7

Do you think that we could do lemmatization before machine translation? Provide 1 argument that for why it might help, and one for why it might make things more complicated. List any assumptions that might make your answer more complicated. (2)

Question 7

Topic: Lecture 8
Source: Lecture 8

Imagine that you're working with a linguist who is not very good with technology. They store all of their data in .docx files, scattered across their desktop. What arguments would you make for them to convert to .tsv or .json, and how would you alleviate their worries that they wouldn't be able to access or modify their information (no, you can't teach them Python)? (2)

Question 8

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 9

Topic: Long
Source: Lecture 5

In class, we've taken a brief look at both prefixes and suffixes, but there are other ways of inflecting words. "circumfixes" wrap around a word, such as the German past participle marker "ge-t" ("ich spiele" - "I play"; "ich habe gespielt" - I have played). Likewise, "infixes" occur inside of a word - "cupful" + Plural -> "cupsful", or in Tagalog: "bili" -> "to buy"; "bumili" -> "X is buying". Finally, "reduplication" occurs when part or all of a token is repeated to indicate some feature, such as repetition or future intent in Tagalog: "aray" -> "day"; "arayaray" -> everyday; "basa" -> "to read"; "babasa" -> "will read (in the future)". Which of these are best suited for regexes, and which features of regexes are they exploiting? Are there any that are mostly unsuited to regexes? Why? (3)

END OF QUIZ