START OF QUIZ
Student ID: 29480787,Komalpreet,

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

Describe the purpose of the various types of brackets in regexes, and how they differ. (1)

Question 2

Topic: Lecture 6
Source: Lecture 6

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

Question 3

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 4

Topic: Lecture 6
Source: Lecture 6

Beautiful Soup parses the children of a tag as a list. Why do you think they didn't use a set, instead, given the faster access times? Give 2 reasons, and briefly explain. (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 8
Source: Lecture 8

In class, I mentioned that we always want to close a file correctly. Beyond freeing up system resources, it also "flushes the buffer", which ensures that any current read or write operations that are in the job queue, but haven't yet been processed, are completed. Knowing what you do about encodings, what is a possible ramification of not flushing the buffer? Explain at least 2. (2)

Question 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 8

Topic: Lecture 5
Source: Lecture 5

Imagine we have a spell-checker that can identify common misspellings of words by replacing certain letters with a capture group that contains letters that are nearby on the keyboard. How aggressive of a regex would we want to write for this (ie, how many letters in the word would we want to replace with a group)? Explain. (2)

Question 9

Topic: Long
Source: Lecture 6

You've been hired by a company that is working with their own version of XML that they call
"NQAXML" (Not-Quite-As-eXtensible Markup Language). It provides stronger restrictions
on tag names (they must be all uppercase, and no longer than 10 characters long), and it
doesn't allow nested spans with identically-named tags. Like HTML, it also has a set of tags
that must appear in every document. Describe your process for creating a data validator
that takes an XML file, and ensures that it satisties the rules of NQAXML. (3)

END OF QUIZ