

START OF QUIZ

Student ID:

30756043, Ting, Lillian

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 8

Source: Lecture 8

Give 2 reasons to use a zip file. (1)

Question 2

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 3

Topic: Lecture 5

Source: Lecture 5

Imagine you have a block of text with paragraphs separated by blank lines. How would you use regex to find the start of each paragraph? What assumptions would you make about the formatting of the text? (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 6

Source: Lecture 6

Suppose you've trained a Named Entity Recognition (NER) model using XML-annotated text data, but it consistently fails to recognize locations. What steps would you take to determine if the problem lies with the model, the training data, or both? What resources would you need to investigate further? (2)

Question 7

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 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 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