START OF QUIZ
Student ID:
47323894,Liao,Jingyi

Question 1

Topic: Lecture 7
Source: Lecture 7

What impact does lemmatization or stemming have on TTR? How might that affect our algorithms? (1)

Question 2

Topic: Lecture 7
Source: Lecture 7

What implications does correct sentence segmentation have on downstream tasks? List at least one assumption we can make if we can assume that our sentences are correctly segmented. (1)

Question 3

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 4

Topic: Lecture 5
Source: Lecture 5

What is the purpose of escaping? (1)

Question 5

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 6

Topic: Lecture 5
Source: Lecture 5

In the last review set, there was a question about identifying valid floats using string operations. How would you do it with a regex? Explain the logic. (1)

Question 7

Topic: Lecture 6
Source: Lecture 6

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

Question 8

Topic: Lecture 6
Source: Lecture 6

In class, we mentioned a few different file types that are actually XML (such as .html, .doc, and .ipynb). Do you think that you could represent a Python library as an XML document? If so, what kind of tags might you need to cover some of the syntactic rules of Python? If not, why not? (2)

Question 9

Topic: Long
Source: Long

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)? (3)

END OF QUIZ