START OF QUIZ
Student ID:
34447581,Ong,Claudia

Question 1

Topic: Lecture 5
Source: Lecture 5

What is the purpose of escaping? (1)

Question 2

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 3

Topic: Lecture 8
Source: Lecture 8

If you were working with an unknown language, which encoding would be most appropriate? Briefly explain. (1)

Question 4

Topic: Lecture 6
Source: Lecture 6

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

Question 5

Topic: Lecture 6
Source: Lecture 6

Imagine that you're building a web scraper, and you find that most of the information presented on the front page is just a collection of links to other pages, so you can't just parse it with an XML parser. What extra functionality would you have to build into your scraper to actually get all the XML data? (2)

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 7
Source: Lecture 7

What impact does lemmatization or stemming have with respect to the Zipfian curve? How might that affect our algorithms? (1)

Question 8

Topic: Lecture 5
Source: Lecture 5

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

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