

START OF QUIZ

Student ID:

67446104,Nguyen,Minh

Question 1

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 2

Topic: Lecture 5

Source: Lecture 5

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

Question 3

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 4

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 5

Topic: Lecture 5

Source: Lecture 5

Write a short function that uses regexes to identify which case (title, camel, or sentence case) a sentence occurs in. (1)

Question 6

Topic: Lecture 7

Source: Lecture 7

Do you think it would be easy or hard to build a training set for a sentence segmenter? Explain. Do you think it would be easier or harder to build a training set for a word tokenizer? What kind of assumptions would you be making about the difficulty of the task? You don't need to worry about the ML tool used - this is about building the training set. (2)

Question 7

Topic: Lecture 6

Source: Lecture 6

What kind of tags do you think would enhance the following sentence? Annotate the sentence with at least two types of tags (they don't have to be accurate - just make sure I can understand the meaning): "Dr. Grant, my dear Dr. Sattler, welcome... to Jurassic Park!" (1)

Question 8

Topic: Lecture 8

Source: Lecture 8

What are 2 benefits of .py files over .ipynb files, and 2 benefits of .ipynb files over .py files? (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