

START OF QUIZ

Student ID:

66605874,Li,Mingcong

Question 1

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 2

Topic: Lecture 5

Source: Lecture 5

In class, I mentioned that we can use regexes to identify typos by finding letters that are close on the keyboard. What functionality of regexes would we exploit to identify these types of characters? Give a brief example. (2)

Question 3

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 4

Topic: Lecture 5

Source: Lecture 5

What is the purpose of escaping? (1)

Question 5

Topic: Lecture 6

Source: Lecture 6

XML can be opened by most plain-text text editors. Name a benefit and a disadvantage of this feature. (1)

Question 6

Topic: Lecture 8

Source: Lecture 8

Why do Python programmers like JSON files so much? (1)

Question 7

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 8

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 9

Topic: Long

Source: Long

Morphological Analysis is a process whereby we recover the lemma and morphologically-informed POS together. For example, the input might be "ran", and the output would be "run + VB;PAST". Do you think it would be best to 1. run tagging first, and then lemmatize using the tag 2. lemmatize first, and then tag, or 3. do both jointly? Why do you think one or the other would be more beneficial, and what information you be leveraging from one to help the other? Do you think this would be harder or easier for inflectionally-rich languages? Justify your answer. As always, state your assumptions. (3)

END OF QUIZ