START OF QUIZ
Student ID: 60226669,Mohammad,Waqas

Question 1

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 2

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 3

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 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

What are 2 benefits of .py files over .ipynb files, and 2 benefits of .ipynb files over .py files? (1)

Question 6

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 7

Imagine that we had a phonetically-transcribed poem (or song). How could we use regexes to identify the rhyme scheme ((since not all of you are familiar with phonetic transcription, you can just describe the logic)? You can assume that each line is written on a new line, and that it is written in stanzas of 4 lines each. List any assumptions. (2)

Question 8

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 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