# START OF QUIZ
# Student ID: 83261909,Dang,Trang

# Question 1

Topic: Lecture 8
Source: Lecture 8

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

# 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 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 4

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 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 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 7

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

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 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