

**START OF QUIZ**

**Student ID:**

**38522413, Wang, Yuxi**

## Question 1

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 2

Topic: Lecture 8

Source: Lecture 8

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

### Question 3

Topic: Lecture 7

Source: Lecture 7

Why is part of speech relevant when lemmatizing? Do you think it would be just as important in stemming? (1)

## Question 4

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 5

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 6

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 7

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 8

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 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**