

**START OF QUIZ**

**Student ID:**

**30821250,Huang,Chloe**

## Question 1

Topic: Lecture 8

Source: Lecture 8

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

## Question 2

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 3

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 4

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 5

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 6

Topic: Lecture 5

Source: Lecture 5

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 7

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 8

Topic: Lecture 8

Source: Lecture 8

If you were working with an unknown language, which encoding would be most appropriate? Briefly explain. (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**