

**START OF QUIZ**  
**Student ID:**  
**29404548, Lee, Ai Ka**

## Question 1

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 2

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 3

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 4

Topic: Lecture 8

Source: Lecture 8

What is the purpose of an archive (2 reasons). (1)

## Question 5

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 6

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 7

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 8

Topic: Lecture 7

Source: Lecture 7

What impact does lemmatization or stemming have on TTR? 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**