

START OF QUIZ

Student ID:

**37105848, Yuen, Shing Hei
Robin**

Question 1

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 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 8

Source: Lecture 8

If you have a corpus of 10 billion words stored in a text file tokenized with one word per line, what is best approach to processing the content of the file after it has been opened? (1)

Question 4

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 5

Topic: Lecture 7

Source: Lecture 7

What impact does lemmatization or stemming have on TTR? How might that affect our algorithms? (1)

Question 6

Topic: Lecture 6

Source: Lecture 6

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 7

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 8

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 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