

START OF QUIZ

Student ID:

33836685,LAM,HUI YIN

Question 1

Topic: Lecture 6

Source: Lecture 6

XML can be opened by most plain-text text editors. Name a benefit and a disadvantage of this feature. (1)

Question 2

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 3

Topic: Lecture 8

Source: Lecture 8

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

Question 4

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 5

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 6

Topic: Lecture 5

Source: Lecture 5

How would you use regexes to do sentence segmentation? Do you think it could correctly identify all cases? Explain. (1)

Question 7

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 8

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 9

Topic: Long

Source: Long

In class, we've taken a brief look at both prefixes and suffixes, but there are other ways of inflecting words. "circumfixes" wrap around a word, such as the German past participle marker "ge-t" ("ich spiele" - "I play"; "ich habe gespielt" - I have played). Likewise, "infixes" occur inside of a word - "cupful" + Plural -> "cupsful", or in Tagalog: "bili" -> "to buy"; "bumili" -> "X is buying". Finally, "reduplication" occurs when part or all of a token is repeated to indicate some feature, such as repetition or future intent in Tagalog: "aray" -> "day"; "arayaray" -> everyday; "basa" -> "to read"; "babasa" -> "will read (in the future)". Which of these are best suited for regexes, and which features of regexes are they exploiting? Are there any that are mostly unsuited to regexes? Why? (3)

END OF QUIZ