

START OF QUIZ

Student ID:

80162985,Sirwani,kartik

Question 1

Topic: Lecture 8

Source: Lecture 8

What are two advantages of using .py files over .ipynb files for deployment, and two reasons why .ipynb files are preferred for prototyping or development? (1)

Question 2

Topic: Lecture 5

Source: Lecture 5

Imagine you have a block of text with paragraphs separated by blank lines. How would you use regex to find the start of each paragraph? What assumptions would you make about the formatting of the text? (1)

Question 3

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 4

Topic: Lecture 7

Source: Lecture 7

In class, we built a POS tagger that tries to give a majority tag to a word; if it's out-of-vocabulary, it backs-off to Regexes. This is clearly overly simplistic. List two assumptions that are being violated by this model. (1)

Question 5

Topic: Lecture 8

Source: Lecture 8

Why should you get into the habit of using "with open()"? Are there any downsides? (1)

Question 6

Topic: Lecture 6

Source: Lecture 6

Suppose you've trained a Named Entity Recognition (NER) model using XML-annotated text data, but it consistently fails to recognize locations. What steps would you take to determine if the problem lies with the model, the training data, or both? What resources would you need to investigate further? (2)

Question 7

Topic: Lecture 7

Source: Lecture 7

Do you think that we could do lemmatization before machine translation? Provide 1 argument that for why it might help, and one for why it might make things more complicated. List any assumptions that might make your answer more complicated. (2)

Question 8

Topic: Lecture 5

Source: Lecture 5

Imagine you are processing a text document where dates are written in multiple formats, such as "12-05-2024", "05/12/2024", or "12 December 2024". How would you write a regex to capture these date formats (just the logic)? What assumptions would you make? (2)

Question 9

Topic: Long

Source: Lecture 6

You've been hired by a company that is working with their own version of XML that they call "NQAXML" (Not-Quite-As-eXtensible Markup Language). It provides stronger restrictions on tag names (they must be all uppercase, and no longer than 10 characters long), and it doesn't allow nested spans with identically-named tags. Like HTML, it also has a set of tags that must appear in every document. Describe your process for creating a data validator that takes an XML file, and ensures that it satisfies the rules of NQAXML. (3)

END OF QUIZ