

**START OF QUIZ**  
**Student ID: 12582557,li**  
**Chen,Mariana**

## Question 1

Topic: Lecture 5

Source: Lecture 5

Write a regex pattern that matches any valid email address (i.e., with basic rules like user@domain.com). What challenges might you face in accurately matching all possible email formats? (1)

## Question 2

Topic: Lecture 7

Source: Lecture 7

Can you think of any classes of words in English where the stem and the lemma will always be identical? Why is that of little interest to us? (1)

### Question 3

Topic: Lecture 8

Source: Lecture 8

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

## Question 4

Topic: Lecture 7

Source: Lecture 7

What is the difference between a stem and a lemma? What impacts does that have on our algorithms? (1)

## Question 5

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 6

Topic: Lecture 8

Source: Lecture 8

In class, I mentioned that we always want to close a file correctly. Beyond freeing up system resources, it also "flushes the buffer", which ensures that any current read or write operations that are in the job queue, but haven't yet been processed, are completed. Knowing what you do about encodings, what is a possible ramification of not flushing the buffer? Explain at least 2. (2)

## Question 7

Topic: Lecture 6

Source: Lecture 6

Consider using XML to represent a machine learning model's architecture. What XML tags might be useful for representing layers, activation functions, and connections between layers (you don't need to describe a deep-learning architecture - describe one you're familiar with)? If this doesn't seem possible, explain why not. (2)



## Question 8

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 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**