#### Mid Term-1 Review

#### Question 1.

Please write a program to count the number of amino acids in the protein file.

## Question 2.

How do you find the spelling errors in the list using the word list corpus. Write a program for it.

### Question 3.

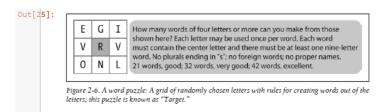
What is the difference between lemmatization and stemming, provide some examples for it.

#### Question 4.

What does the conditional frequency distribution plot show. Please explain it with a relevant program.

## Question 5.

Write a program for the following questions with the following scenario.



Our program iterates through every word and, for each one, checks whether it meets the conditions. It is easy to check obligatory letter and length constraints (and we'll only look for words with six or more letters here). It is trickier to check that candidate solutions only use combinations of the supplied letters, especially since some of the supplied letters appear twice (here, the letter v). The FreqDist comparison method permits us to check that the frequency of each letter in the candidate word is less than or equal to the frequency of the corresponding letter in the puzzle.

## Question 6.

Write a program to extract. Mr. Simpson, Mr. Brown and Mr. T from the following string.

hello world

1223

2020-05-20

Mr Simpson

Ars Simpson
⁄lr. Brown
As Smith
Лr. T
ni <i>n</i>
Question 7
Vhat is the objective of Natural Language Processing. Can you explain natural language processing ipeline.
Question 8
xplain concatenation, flatten, and argmax functions in the context of NumPy array? Why do you nee lumPy array?
Question 9
Vrite a program to display a unique element in the list, do not use set data structure? (5)
Question 10
Vrite a program to merge two data frames based on a unique column?
Question 11
Vhat is One Hot vector and Document Term matrix.
Question 12
What is TF-IDF, construct a TFID feature vector for the following four documents.
D1 People watch Campusx
D2 Campusy watch Campusy

а

D1	People watch Campusx
D2	Campusx watch Campusx
D3	People write Comment
D4	Campusx write Comment

https://youtu.be/2FeDYJZaeU8

# Question 13:

# Explain

- a) Lexical diversity
- b) Hypernyms, hyponyms
- c) Polysemy
- d) "Today is the midterm test 1" tokenize the string based upon the space.
- e) What is one hot vector list out its advantages and disadvantages.

# Question 14:

Explain inheritance in object-oriented programming (OOP) with super() function in it, can you write a program for it. What are getters and setters in the context of OOP?

# Question 15:

Can you explain the dot product concept in the NLP and why is it used with NLP program?