Handling Text

```
s = "A simple given sentence"
t = "simple"
u = "complex"
s.find(t)
#index of first instance of string t inside s (-1 if not found)
s.rfind(t)
#index of last instance of string t inside s (-1 if not found)
s.index(t)
# like s.find(t) except it raises ValueError if not found
s.rindex(t)
#like s.rfind(t) except it raises ValueError if not found
s.join(t)
#combine the words of the text into a string using s as the glue
s.split(t)
#split s into a list wherever a t is found (whitespace by default)
s.splitlines()
#split s into a list of strings, one per line
s.lower()
#a lowercase version of the string s
s.upper()
#an uppercased version of the string s
s.title()
#a titlecased version of the string s
s.strip()
#a copy of s without leading or trailing whitespace
s.replace(t, u)
#replace instances of t with u inside s
```

1. Given the following sentence for the following question:

   Sentence = “I am learning Natural Language Processing”

   a. Capitalize the first words of each word of the sentence (Let’s start simple!)

   b. Replace the word “learning” with “exploring”

   c. Is it possible to abbreviate the word the Natural Language Processing as simply NLP with all caps? If so, can you code in doing that

```
#substrings
var="I am learning NLP"
f= "learn"
var.find(f)
```

Normalization and Processing Data

2. Given the following text,

   text=['This is introduction to NLP','It is likely to be useful,

   to people ','Machine learning is the new electricity',

   'There would be less hype around AI and more action going

   forward','python is the best tool!','R is good langauage',

   'I. like. This book!','I want more books like this']

   a. Can you remove the punctuations? (Consider the one before the last sentence),

      Hint: Use a simple regex expression

   b. As in Question Number 1, we wanted to make an association between abbreviations like NLP and Natural Language processing. Make a **look up dictionary** of the following words

      NLP = Natural Language Processing

      ML - Machine Learning

AI - Artificial Intelligence

c. Now, according to the look_up dictionary, write a function that can standardize given an input text. You can either convert all to abbreviations, or you can expand the abbreviation.
d. Consider the third sentence from the given text and change it into = “Mahine laering is the new electrcity”. Can you make spelling amends for words that are spelled wrong? Write a simple code to make changes. Also, can we generalize it for other conditions?

Tokenization

3. Given a text

   Text = “NLP is very interesting”

   a. Can you use a simple tokenizing built in function split() to tokenize the words from the text? Write a module to tokenize texts
   b. Text = “NLP is very interesting. It is interesting because it gives an understanding on how machines will learn to understand human spoken language”. If we modify the sentences, can you do a simple tokenizations again
   c. Given sentence in b, can you count the frequency of words and print out the association.

Stemming and Lemmatization

4. Consider the following text:

   Text = ['I like fishing','I eat fish','There are many fishes in Pound' , ‘leaves and leaf’]

   a. Write a code to stem the word “fish”
   b. Can we generalize it into other conditions?
   c. Write a code to lemmatize the word “leaf”