



Department of Computer Science & Engineering  
University College of Engineering, JNT University Kakinada

II B. Tech II Seme-R20

PYTHON PROGRAMMING: NLP  
Code: R20SO2201

**Pre-requisites:**

- Basic Knowledge of Natural Language Processing
- Hands-on practice of Python
- Basic idea of NLP using the library- SpaCy

**Experiment-1:**

- a) Getting started with NLTK, install NLTK using PIP
- b) Try using the Python interpreter as a calculator, and typing expressions like  $12 / (4 + 1)$ .

**Experiment-2:**

- a) Define a string and assign it to a variable, e.g., `my_string = 'My String'` (but put something more interesting in the string). Print the contents of this variable in two ways, first by simply typing the variable name and pressing enter, then by using the print statement.
- b) Try adding the string to itself using `my_string + my_string`, or multiplying it by a number, e.g., `my_string * 3`. Notice that the strings are joined together without any spaces. How could you fix this?

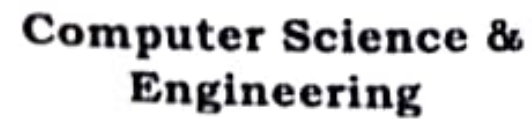
**Experiment-3:** Define a variable `my_sent` to be a list of words, using the syntax `my_sent = ["My", "sent"]` (but with your own words, or a favorite saying).

- a. Use `' '.join(my_sent)` to convert this into a string.
- b. Use `split()` to split the string back into the list form you had to start with.

**Experiment-4:** Write expressions for finding all words in `text6` that meet the conditions listed below. The result should be in the form of a list of words: `['word1', 'word2', ...]`.

- a. Ending in *ise*
- b. Containing the letter *z*
- c. Containing the sequence of letters *pt*
- d. Having all lowercase letters except for an initial capital (i.e., titlecase)





**Experiment-5:** Define sent to be the list of words ['she', 'sells', 'sea', 'shells', 'by', 'the', 'sea', 'shore']. Now write code to perform the following tasks:

- Experiment-6:** Write a program by using list addition, and the set and sorted operations, compute the vocabulary of the sentences sent1 ... sent8.

**Experiment-8:** Write a program for defining a function percent (word, text) that calculates how often a given word occurs in a text, and expresses the result as a percentage.

**Experiment-10:** Write a program to find all the four-letter words in the Chat Corpus (text5). With the help of a frequency distribution (FreqDist), show these words in decreasing order of frequency.

**Experiment-12:** Tom is a comic editor in a X company one day while editing a particular script he became enthusiastic about the story of the script so there is no time to read the whole script he decided to understand the total story line by learning about the characters so he wants to separate the words in the sentence to know the characters as the whole story is complex. So implement a python program that splits the words and display both splitted words and count of the words in the given sentence using tokenizer function?

1. <https://www.nltk.org/book/ch01.html>
2. <https://www.google.com/books/edition/nltk-manual>