

# School of Computer Science Engineering and Technology

Course- B. Tech  
Course Code- CSET346

Year- 2022  
Date: 23-08-2022

Type- Elective  
Course Name: Natural language  
processing  
Semester- odd  
Batch- ALL

## Lab Assignment 02 – Second Step towards data pre-processing

The main objective of this assignment is to know pre-processing of text data through spacy library.

You will learn how to achieve sentence segmentation, phrase matching and vocabulary by incorporating spacy and its modules.

### 1. Sentence segmentation

Dividing a paragraph based on sentences. This process is known as Sentence Segmentation. In Python, we implement this part of NLP using the spacy library.

Q1. Choose any text data and apply Segmentation using spacy library.

### 2. Phrase matching

The spaCy library comes with Matcher tool that can be used to specify custom rules for phrase matching.

Q2. Apply phrase matching for the following data,

**"Glowing review overall, and some really interesting side-by-side "**

**"Photography tests pitting the iPhone 11 Pro against the "**

**"Galaxy Note 10 Plus and last year's iPhone XS and Google Pixel 3."**

### 3. nltk vs spacy

Q3. Analyse the advantage of spacy over nltk on the basis time of execution.