**ASSIGNMENT - 1**

**NATURAL LANGUAGE PROCESSING**

**PRIYANSHU SHARMA**

**15BCE1282**

**POS TAGGING**

**CODE**

import nltk

from nltk.corpus import state\_union

from nltk.tokenize import PunktSentenceTokenizer

train\_text = state\_union.raw("2005-GWBush.txt")

sample\_text = state\_union.raw("2006-GWBush.txt")

custom\_sent\_tokenizer = PunktSentenceTokenizer(train\_text)

tokenized = custom\_sent\_tokenizer.tokenize(sample\_text)

def process\_content():

try:

for i in tokenized:

words = nltk.word\_tokenize(i)

tagged = nltk.pos\_tag(words)

print(tagged)

except Exception as e:

print(str(e))

process\_content()

**OUTPUT**

****

****

**SENTENCE SEGMENTATION**

**CODE**

from nltk.tokenize import sent\_tokenize, word\_tokenize

example\_text = "Hello Mr. Priyanshu! How are you? My first paper is on Monday, and I am really happy."

print(sent\_tokenize(example\_text)) #sentence segmentation

print(word\_tokenize(example\_text)) #word by word segmentation

for i in word\_tokenize(example\_text): #print all the segmenated word

print(i)

**OUTPUT**





