```python
# -*- coding: utf-8 -*-
"""Assignment 7.ipynb

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/1x2YXT6PakGhOrvcTGw-
KgX55sZ3r3Kx4
"""

# Step1:Download the required packages
import nltk
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')

#Step 2:Initialize the text
text= "Tokenization is the first step in text analytics. The process of
breaking down a text paragraph into smaller chunks such as words or
sentences is called Tokenization."

#Step 3:Perform Tokenization
from nltk.tokenize import sent_tokenize
tokenized_text= sent_tokenize(text)
print(tokenized_text)

#Step4:Removing Punctuations and Stop Word
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

example_sent = """This is a sample sentence,
                  showing off the stop words filtration."""

stop_words = set(stopwords.words('english'))

word_tokens = word_tokenize(example_sent)
# converts the words in word_tokens to lower case and then checks whether
#they are present in stop_words or not
filtered_sentence = [w for w in word_tokens if not w.lower() in
stop_words]
#with no lower case conversion
filtered_sentence = []

for w in word_tokens:
    if w not in stop_words:
        filtered_sentence.append(w)

print(word_tokens)
print(filtered_sentence)

#Step 6:Perform Stemming
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize

ps = PorterStemmer()

# choose some words to be stemmed
```

```
words = ["program", "programs", "programmer", "programming",
"programmers"]

for w in words:
    print(w, " : ", ps.stem(w))

Step 6:Perform Stemming
from nltk.stem import PorterStemmer
e_words= ["wait", "waiting", "waited", "waits"]
ps =PorterStemmer()
for w in e_words:
  rootWord=ps.stem(w)
  print(rootWord)

#step 7:Perform Lemmatization
from nltk.stem import WordNetLemmatizer
wordnet_lemmatizer = WordNetLemmatizer()
text = "studies studying cries cry"
tokenization = nltk.word_tokenize(text)
for w in tokenization:
  print("Lemma for {} is {}".format(w,
  wordnet_lemmatizer.lemmatize(w)))

#Step 8:Apply POS Tagging to text
import nltk
from nltk.tokenize import word_tokenize
data="The pink sweater fit her perfectly"
words=word_tokenize(data)
for word in words:
  print(nltk.pos_tag([word]))
```