# Experiment No.: 11

## Aim

Natural Language Processing

Problems may be designed for the following topics so that students can get hands on experience in using python for natural language processing:

• Part of Speech tagging

• N-gram and smoothening

• Chunking

**CO5**

Implement programs for web data mining and natural language processing using NLTK

## Procedure

### Tagging

```
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize, sent_tokenize
stop_words = set(stopwords.words('english'))
txt = "Sukanya, Rajib and Naba are my good friends. " \
      "Sukanya is getting married next year. " \
      "Marriage is a big step in one’s life." \
      "It is both exciting and frightening. " \
      "But friendship is a sacred bond between people." \
      "It is a special kind of love between us. " \
      "Many of you must have tried searching for a friend " \
      "but never found the right one."
tokenized = sent_tokenize(txt)
for i in tokenized:
    wordsList = nltk.word_tokenize(i)
    wordsList = [w for w in wordsList if not w in stop_words]
```

```
tagged = nltk.pos_tag(wordsList)
print(tagged)
```

**N-gram**

```
import nltk
# nltk.download()
from nltk.util import ngrams
samplText = 'welcome to amal jyothi college of engineering'
NGRAMS = ngrams(sequence=nltk.word_tokenize(samplText), n=3)
for grams in NGRAMS:
print(grams)
```

**Chunking**

```
import nltk
new = "The big cat ate the little mouse who was after the fresh cheese"
new_tokens = nltk.word_tokenize(new)
print(new_tokens)
new_tag = nltk.pos_tag(new_tokens)
print(new_tag)
grammer = "NP: {<DT>?<JJ>*<NN>}"
chunkParser = nltk.RegexpParser(grammer)
chunked = chunkParser.parse(new_tag)
print(chunked)
chunked.draw()
```

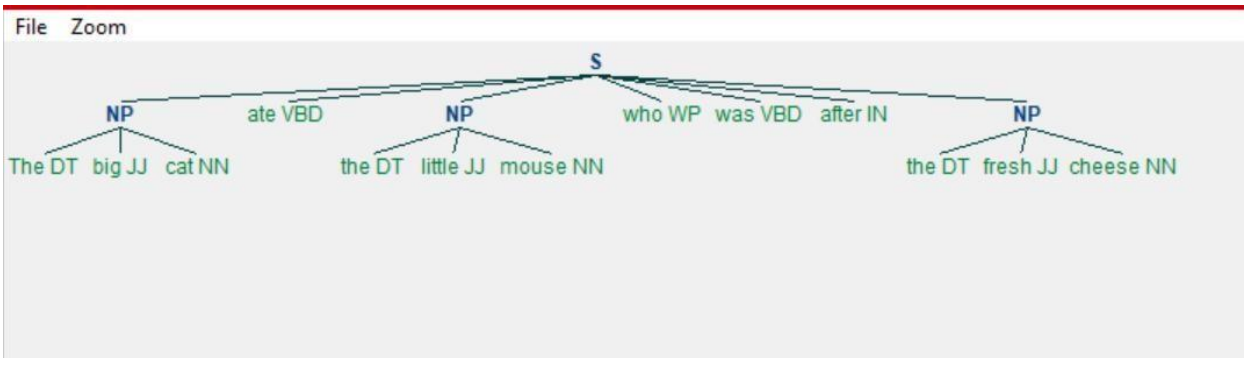
## Output Screenshot

```
prgm3
C:\Users\ajcemca\PycharmProjects\pythonProject0411\venv\Scripts\python.exe C:/Users/ajcemca/PycharmProjects/pythonProje
[('Sukanya', 'NNP'), (',', ','), ('Rajib', 'NNP'), ('Naba', 'NNP'), ('good', 'JJ'), ('friends', 'NNS'), ('.', '.')]
[('Sukanya', 'NNP'), ('getting', 'VBG'), ('married', 'VBN'), ('next', 'JJ'), ('year', 'NN'), ('.', '.')]
[('Marriage', 'NN'), ('big', 'JJ'), ('step', 'NN'), ('one', 'CD'), ('’', 'NN'), ('life.It', 'NN'), ('exciting', 'VBG'),
[('But', 'CC'), ('friendship', 'NN'), ('sacred', 'VBD'), ('bond', 'NN'), ('people.It', 'NN'), ('special', 'JJ'), ('kind
[('Many', 'JJ'), ('must', 'MD'), ('tried', 'VB'), ('searching', 'VBG'), ('friend', 'NN'), ('never', 'RB'), ('found', 'V

Process finished with exit code 0
```

```
prgm2
C:\Users\ajcemca\PycharmProjects\pythonProject0411\venv\Scripts\python.exe
('welcome', 'to', 'amal')
('to', 'amal', 'jyothi')
('amal', 'jyothi', 'college')
('jyothi', 'college', 'of')
('college', 'of', 'engineering')

Process finished with exit code 0
```




## Result

The program was executed and the result was successfully obtained. Thus CO5 was obtained.