In [4]:
```
# Sample code to remove noisy words from a text

noise_list = ["is", "a", "this", "..."]
def _remove_noise(input_text):
    words = input_text.split()
    noise_free_words = [word for word in words if word not in noise_list]
    noise_free_text = " ".join(noise_free_words)
    return noise_free_text

_remove_noise("Sharmila k is a Faculty of NLP and I Suni is the Student")
```

Out[4]:
```
'Sharmila k Faculty of NLP and I Suni the Student'
```

In [3]:
```
# Sample code to remove a regex pattern
import re

def _remove_regex(input_text, regex_pattern):
    urls = re.finditer(regex_pattern, input_text)
    for i in urls:
        input_text = re.sub(i.group().strip(), '', input_text)
    return input_text

regex_pattern = "#[\w]*"
_remove_regex("remove this #hashtag from analytics Sunil and Sharmila bhanu K", rege
```

Out[3]:
```
'remove this  from analytics Sunil and Sharmila bhanu K'
```

In [17]:
```
from nltk.stem.wordnet import WordNetLemmatizer
lem = WordNetLemmatizer()

from nltk.stem.porter import PorterStemmer
stem = PorterStemmer()

word = "multiplying"
lem.lemmatize(word, "v")
```

Out[17]:
```
'multiply'
```

In [18]:
```
stem.stem(word)
```

Out[18]:
```
'multipli'
```

In [35]:
```
def generate_ngrams(text, n):
    text='this is my name,2'
    words = text.split()
    output = []
    for i in range(len(words)-n+1):
        output.append(words[i:i+n])
    return output
```

In [36]:
```
generate_ngrams('this is a sample text', 2)
```

Out[36]:
```
[['this', 'is'], ['is', 'my'], ['my', 'name,2']]
```

In [0]: