INSTALLING NLTK

In [4]:
```
!pip install nltk
nltk.download('punkt')
```

```
Requirement already satisfied: nltk in c:\users\abhiram\anaconda3\lib\site-pack
ages (3.6.5)
Requirement already satisfied: click in c:\users\abhiram\anaconda3\lib\site-pac
kages (from nltk) (8.0.3)
Requirement already satisfied: joblib in c:\users\abhiram\anaconda3\lib\site-pa
ckages (from nltk) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\abhiram\anaconda3\li
b\site-packages (from nltk) (2021.8.3)
Requirement already satisfied: tqdm in c:\users\abhiram\anaconda3\lib\site-pack
ages (from nltk) (4.62.3)
Requirement already satisfied: colorama in c:\users\abhiram\anaconda3\lib\site-
packages (from click->nltk) (0.4.4)

[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\ABHIRAM\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping tokenizers\punkt.zip.
```

Out[4]: True

Importing required libraries

In [2]:
```
import nltk
from nltk.tokenize import word_tokenize, sent_tokenize
from nltk.corpus import RegexpTokenizer as regextoken

nltk.download('stopwords')

nltk.download('wordnet')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\ABHIRAM\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\ABHIRAM\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

Out[2]: True

Tokenzing

In [3]:
```python
str1='{"orders":[{"id":1},{"id":2},{"id":3},{"id":4},{"id":5},{"id":6},{"id":7},{
token= word_tokenize(str1)
print(token)
```

```
['{', '``', 'orders', "'", ':', '[', '{', '``', 'id', "'", ':1', '}', ',',
'{', '``', 'id', "'", ':2', '}', ',', '{', '``', 'id', "'", ':3', '}', ',',
'{', '``', 'id', "'", ':4', '}', ',', '{', '``', 'id', "'", ':5', '}', ',',
'{', '``', 'id', "'", ':6', '}', ',', '{', '``', 'id', "'", ':7', '}', ',',
'{', '``', 'id', "'", ':8', '}', ',', '{', '``', 'id', "'", ':9', '}', ',',
'{', '``', 'id', "'", ':10', '}', ',', '{', '``', 'id', "'", ':11', '}', ',',
'{', '``', 'id', "'", ':648', '}', ',', '{', '``', 'id', "'", ':649', '}',
',', '{', '``', 'id', "'", ':650', '}', ',', '{', '``', 'id', "'", ':651',
'}', ',', '{', '``', 'id', "'", ':652', '}', ',', '{', '``', 'id', "'", ':65
3', '}', ']', ',', "'", 'errors', "'", ':', '[', '{', '``', 'code', "'", ':
3', ',', "'", 'message', "'", ':', "'", '[', 'PHP', 'Warning', '#', '2',
']', 'count', '(', ')', ':', 'Parameter', 'must', 'be', 'an', 'array', 'or', 'a
n', 'object', 'that', 'implements', 'Countable', '(', '153', ')', "'", '}',
']', '}']
```

Using Regex extrating Required output

In [4]:
```python
import re
[w for w in token if re.search('^:[0-9]\d*', w)]
```

Out[4]:
```
[':1',
 ':2',
 ':3',
 ':4',
 ':5',
 ':6',
 ':7',
 ':8',
 ':9',
 ':10',
 ':11',
 ':648',
 ':649',
 ':650',
 ':651',
 ':652',
 ':653',
 ':3']
```