Name: Harsh Zanwar Roll No:73 Branch:CSE(AIML)

In [10]:
```
pip install nltk
```

Requirement already satisfied: nltk in c:\users\asus\appdata\local\programs\python\python310\lib\site-packages (3.7)
Requirement already satisfied: joblib in c:\users\asus\appdata\local\programs\python\python310\lib\site-packages (from nltk) (1.2.0)
Requirement already satisfied: click in c:\users\asus\appdata\local\programs\python\python310\lib\site-packages (from nltk) (8.1.3)
Requirement already satisfied: tqdm in c:\users\asus\appdata\local\programs\python\python310\lib\site-packages (from nltk) (4.64.1)
Requirement already satisfied: regex>=2021.8.3 in c:\users\asus\appdata\local\programs\python\python310\lib\site-packages (from nltk) (2022.10.31)
Requirement already satisfied: colorama in c:\users\asus\appdata\local\programs\python\python310\lib\site-packages (from click->nltk) (0.4.5)
Note: you may need to restart the kernel to use updated packages.

WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local\programs\python\python310\lib\site-packages)

[notice] A new release of pip is available: 23.0.1 -> 23.1.1
[notice] To update, run: python.exe -m pip install --upgrade pip

In [11]:
```
import nltk
```

In [3]:
```
nltk.download()
```

showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml (https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml)

Out[3]: True

In [4]:
```python
from nltk.book import *
```

```
*** Introductory Examples for the NLTK Book ***
Loading text1, ..., text9 and sent1, ..., sent9
Type the name of the text or sentence to view it.
Type: 'texts()' or 'sents()' to list the materials.
text1: Moby Dick by Herman Melville 1851
text2: Sense and Sensibility by Jane Austen 1811
text3: The Book of Genesis
text4: Inaugural Address Corpus
text5: Chat Corpus
text6: Monty Python and the Holy Grail
text7: Wall Street Journal
text8: Personals Corpus
text9: The Man Who Was Thursday by G . K . Chesterton 1908
```

In [5]:
```python
text1.concordance("monstrous")
```

```
Displaying 11 of 11 matches:
ong the former , one was of a most monstrous size . ... This came towards us
,
ON OF THE PSALMS . " Touching that monstrous bulk of the whale or ork we have
r
ll over with a heathenish array of monstrous clubs and spears . Some were thi
ck
d as you gazed , and wondered what monstrous cannibal and savage could ever h
av
that has survived the flood ; most monstrous and most mountainous ! That Himm
al
they might scout at Moby Dick as a monstrous fable , or still worse and more
de
th of Radney .'" CHAPTER 55 Of the Monstrous Pictures of Whales . I shall ere
l
ing Scenes . In connexion with the monstrous pictures of whales , I am strong
ly
ere to enter upon those still more monstrous stories of them which are to be
fo
ght have been rummaged out of this monstrous cabinet there is no telling . Bu
t
of Whale - Bones ; for Whales of a monstrous size are oftentimes cast up dead
u
```

Type *Markdown* and LaTeX: α^2

In [6]:
```python
sent1=["Rehan","shantanu","Danish","."]
len(sent1)
```

Out[6]: 4

In [7]:
```python
def lexical_diversity(text):
    return len(set(text))/len(text)
```

```
In [8]: lexical_diversity("Harsh")
```

```
Out[8]: 1.0
```

```
In [9]: lexical_diversity(text3)
```

```
Out[9]: 0.06230453042623537
```

```
In [10]: f=open('nplsample.txt')
```

```
In [11]: raw=f.read()
```

```
In [12]: raw
```

```
Out[12]: 'The history of Java is very interesting. Java was originally designed for in
         teractive television, but it was too advanced technology for the digital cabl
         e television industry at the time. The history of Java starts with the Green
         Team. Java team members (also known as Green Team), initiated this project to
         develop a language for digital devices such as set-top boxes, televisions, et
         c. However, it was best suited for internet programming. Later, Java technolo
         gy was incorporated by Netscape.'
```

```
In [18]: import pandas as pd
```

```
In [19]: f=pd.read_csv('whetherdata.csv.csv')
         f
```

Out[19]:

	Time	Weather	Temperature	Company	Humidity	Wind	Goes
0	Morning	Sunny	Warm	Yes	Mild	Strong	Yes
1	Evening	Rainy	Cold	No	Mild	Normal	No
2	Morning	Sunny	Moderate	Yes	Normal	Normal	Yes
3	Evening	Sunny	Cold	Yes	High	Strong	Yes

```
In [21]: pip install urllib2
```

```
Note: you may need to restart the kernel to use updated packages.

WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
ERROR: Could not find a version that satisfies the requirement urllib2 (from
versions: none)
ERROR: No matching distribution found for urllib2
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
```

In [22]:
```
pip install --upgrade pip
```

Requirement already satisfied: pip in c:\users\asus\appdata\local\programs\py
thon\python310\lib\site-packages (23.0.1)
Note: you may need to restart the kernel to use updated packages.

WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)

In [23]:
```
nltk.download('punkt')
```

[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\ASUS\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!

Out[23]: True

In [12]:
```
from nltk.tokenize import word_tokenize
```

In [26]:
```
text="Hello my name is Harsh"
word_tokenize(text)
```

Out[26]: ['Hello', 'my', 'name', 'is', 'Harsh']

```python
In [27]:  from bs4 import BeautifulSoup

          # Importing the HTTP library
          import requests as req

          # Requesting for the website
          Web = req.get('https://festive-knuth-1279a2.netlify.app/')

          # Creating a BeautifulSoup object and specifying the parser
          S = BeautifulSoup(Web.text, 'lxml')

          # Using the prettify method
          print(S.prettify())
```

```html
<!DOCTYPE html>
<html lang="en">
 <head>
  <meta charset="utf-8"/>
  <meta content="width=device-width, initial-scale=1.0" name="viewport"/>
  <meta content="ie=edge" http-equiv="X-UA-Compatible"/>
  <link href="./main.css" rel="stylesheet"/>
  <title>
   Document
  </title>
 </head>
 <body>
  <div id="container">
   <div class="form-wrap">
    <h1>
     Sign Up
    </h1>
    <p>
     It's free and only takes a minute
    </p>
    <form>
     <div class="form-group">
      <label for="first-name">
       First Name
      </label>
      <input id="first-name" name="First Name" type="text"/>
     </div>
     <div class="form-group">
      <label for="last-name">
       Last Name
      </label>
      <input id="last-name" name="Last Name" type="text"/>
     </div>
     <div class="form-group">
      <label for="email">
       Email
      </label>
      <input id="email" name="Email" type="text"/>
     </div>
     <div class="form-group">
      <label for="password">
       Password
      </label>
      <input id="password" name="Password" type="password"/>
     </div>
     <div class="form-group">
      <label for="password2">
       Confirm Password
      </label>
      <input id="password2" name="password2" type="password"/>
     </div>
     <button class="btn" type="submit">
      Sign Up
     </button>
     <p class="bottom-text">
      By clicking the Sign Up button, you agree to our
      <a href="#">
```

```
      Terms &amp; Conditions
     </a>
     and
     <a href="#">
      Privacy Policy
     </a>
    </p>
   </form>
  </div>
  <footer>
   <p>
    Already have an account?
    <a href="#">
     Login Here
    </a>
   </p>
  </footer>
 </div>
 </body>
</html>
```

In [28]:
```python
Str="Hello , How are you ?"
token1=word_tokenize(Str)
len(token1)
```

Out[28]: 6

In [29]:
```
pip install spacy
```

```
Requirement already satisfied: spacy in c:\users\asus\appdata\local\programs
\python\python310\lib\site-packages (3.4.3)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in c:\users\asus\appda
ta\local\programs\python\python310\lib\site-packages (from spacy) (2.0.8)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\asus\app
data\local\programs\python\python310\lib\site-packages (from spacy) (1.0.9)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in c:\users\asus\appdata\l
ocal\programs\python\python310\lib\site-packages (from spacy) (0.7.0)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.10 in c:\users\asus\a
ppdata\local\programs\python\python310\lib\site-packages (from spacy) (3.0.1
0)
Requirement already satisfied: packaging>=20.0 in c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages (from spacy) (21.3)
Requirement already satisfied: pathy>=0.3.5 in c:\users\asus\appdata\local\pr
ograms\python\python310\lib\site-packages (from spacy) (0.6.2)
Requirement already satisfied: wasabi<1.1.0,>=0.9.1 in c:\users\asus\appdata
\local\programs\python\python310\lib\site-packages (from spacy) (0.10.1)
Requirement already satisfied: setuptools in c:\users\asus\appdata\local\prog
rams\python\python310\lib\site-packages (from spacy) (58.1.0)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\asus\appdata\l
ocal\programs\python\python310\lib\site-packages (from spacy) (2.0.7)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in c:\users\asus\a
ppdata\local\programs\python\python310\lib\site-packages (from spacy) (1.0.3)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\asus\appdata\l
ocal\programs\python\python310\lib\site-packages (from spacy) (4.64.1)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\asus\appdata
\local\programs\python\python310\lib\site-packages (from spacy) (3.0.8)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in c:\users\asus\appdata\l
ocal\programs\python\python310\lib\site-packages (from spacy) (2.4.5)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\asus\appda
ta\local\programs\python\python310\lib\site-packages (from spacy) (2.28.1)
Requirement already satisfied: thinc<8.2.0,>=8.1.0 in c:\users\asus\appdata\l
ocal\programs\python\python310\lib\site-packages (from spacy) (8.1.5)
Requirement already satisfied: numpy>=1.15.0 in c:\users\asus\appdata\local\p
rograms\python\python310\lib\site-packages (from spacy) (1.23.2)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in c:\users\asus\appda
ta\local\programs\python\python310\lib\site-packages (from spacy) (3.3.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4 in c:\us
ers\asus\appdata\local\programs\python\python310\lib\site-packages (from spac
y) (1.10.2)
Requirement already satisfied: jinja2 in c:\users\asus\appdata\local\programs
\python\python310\lib\site-packages (from spacy) (3.1.2)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\asus\appd
ata\local\programs\python\python310\lib\site-packages (from packaging>=20.0->
spacy) (3.0.9)
Requirement already satisfied: smart-open<6.0.0,>=5.2.1 in c:\users\asus\appd
ata\local\programs\python\python310\lib\site-packages (from pathy>=0.3.5->spa
cy) (5.2.1)
Requirement already satisfied: typing-extensions>=4.1.0 in c:\users\asus\appd
ata\local\programs\python\python310\lib\site-packages (from pydantic!=1.8,!=
1.8.1,<1.11.0,>=1.7.4->spacy) (4.4.0)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\asus\appd
ata\local\programs\python\python310\lib\site-packages (from requests<3.0.0,>=
2.13.0->spacy) (2.1.1)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\asus\appdata
\local\programs\python\python310\lib\site-packages (from requests<3.0.0,>=2.1
3.0->spacy) (1.26.12)
```

```
Requirement already satisfied: idna<4,>=2.5 in c:\users\asus\appdata\local\pr
ograms\python\python310\lib\site-packages (from requests<3.0.0,>=2.13.0->spac
y) (3.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\asus\appdata\lo
cal\programs\python\python310\lib\site-packages (from requests<3.0.0,>=2.13.0
->spacy) (2022.9.24)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in c:\users\asus\appd
ata\local\programs\python\python310\lib\site-packages (from thinc<8.2.0,>=8.
1.0->spacy) (0.0.3)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in c:\users\asus\appdata\lo
cal\programs\python\python310\lib\site-packages (from thinc<8.2.0,>=8.1.0->sp
acy) (0.7.9)
Requirement already satisfied: colorama in c:\users\asus\appdata\local\progra
ms\python\python310\lib\site-packages (from tqdm<5.0.0,>=4.38.0->spacy) (0.4.
5)
Requirement already satisfied: click<9.0.0,>=7.1.1 in c:\users\asus\appdata\l
ocal\programs\python\python310\lib\site-packages (from typer<0.8.0,>=0.3.0->s
pacy) (8.1.3)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages (from jinja2->spacy) (2.1.1)
Note: you may need to restart the kernel to use updated packages.

WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
```

In [30]: `pip install gensim`

```
Requirement already satisfied: gensim in c:\users\asus\appdata\local\programs
\python\python310\lib\site-packages (4.3.0)
Requirement already satisfied: Cython==0.29.32 in c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages (from gensim) (0.29.32)
Requirement already satisfied: scipy>=1.7.0 in c:\users\asus\appdata\local\pr
ograms\python\python310\lib\site-packages (from gensim) (1.9.3)
Requirement already satisfied: smart-open>=1.8.1 in c:\users\asus\appdata\loc
al\programs\python\python310\lib\site-packages (from gensim) (5.2.1)
Requirement already satisfied: FuzzyTM>=0.4.0 in c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages (from gensim) (2.0.5)
Requirement already satisfied: numpy>=1.18.5 in c:\users\asus\appdata\local\p
rograms\python\python310\lib\site-packages (from gensim) (1.23.2)
Requirement already satisfied: pyfume in c:\users\asus\appdata\local\programs
\python\python310\lib\site-packages (from FuzzyTM>=0.4.0->gensim) (0.2.25)
Requirement already satisfied: pandas in c:\users\asus\appdata\local\programs
\python\python310\lib\site-packages (from FuzzyTM>=0.4.0->gensim) (1.4.3)
Requirement already satisfied: pytz>=2020.1 in c:\users\asus\appdata\local\pr
ograms\python\python310\lib\site-packages (from pandas->FuzzyTM>=0.4.0->gensi
m) (2022.2.1)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\asus\appdat
a\local\programs\python\python310\lib\site-packages (from pandas->FuzzyTM>=0.
4.0->gensim) (2.8.2)
Requirement already satisfied: fst-pso in c:\users\asus\appdata\local\program
s\python\python310\lib\site-packages (from pyfume->FuzzyTM>=0.4.0->gensim)
(1.8.1)
Requirement already satisfied: simpful in c:\users\asus\appdata\local\program
s\python\python310\lib\site-packages (from pyfume->FuzzyTM>=0.4.0->gensim)
(2.9.0)
Requirement already satisfied: six>=1.5 in c:\users\asus\appdata\local\progra
ms\python\python310\lib\site-packages (from python-dateutil>=2.8.1->pandas->F
uzzyTM>=0.4.0->gensim) (1.16.0)
Requirement already satisfied: miniful in c:\users\asus\appdata\local\program
s\python\python310\lib\site-packages (from fst-pso->pyfume->FuzzyTM>=0.4.0->g
ensim) (0.0.6)
Requirement already satisfied: requests in c:\users\asus\appdata\local\progra
ms\python\python310\lib\site-packages (from simpful->pyfume->FuzzyTM>=0.4.0->
gensim) (2.28.1)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\asus\appdata\lo
cal\programs\python\python310\lib\site-packages (from requests->simpful->pyfu
me->FuzzyTM>=0.4.0->gensim) (2022.9.24)
Requirement already satisfied: idna<4,>=2.5 in c:\users\asus\appdata\local\pr
ograms\python\python310\lib\site-packages (from requests->simpful->pyfume->Fu
zzyTM>=0.4.0->gensim) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\asus\appdata
\local\programs\python\python310\lib\site-packages (from requests->simpful->p
yfume->FuzzyTM>=0.4.0->gensim) (1.26.12)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\asus\appd
ata\local\programs\python\python310\lib\site-packages (from requests->simpful
->pyfume->FuzzyTM>=0.4.0->gensim) (2.1.1)
Note: you may need to restart the kernel to use updated packages.
```

```
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
WARNING: Ignoring invalid distribution -treamlit (c:\users\asus\appdata\local
\programs\python\python310\lib\site-packages)
```

In [1]:
```python
import spacy
```

In [3]:
```python
spacy_nlp_obj = spacy.load('en_core_web_sm')

test_sentence_4 = 'NLP is a very powerful tool'
test_doc = spacy_nlp_obj(test_sentence_4)
```

In [5]:
```python
random_text = 'It was amazing! It is a great pleasure to be a part of the band.
random_text_doc = spacy_nlp_obj(random_text)
print("Text".ljust(10), ' ', "Alpha", "Space", "Stop", "Punct")
for token in random_text_doc:
    print(token.text.ljust(10), ':', token.is_alpha, token.is_space, token.is_s
```

```
Text        Alpha Space Stop Punct
It         : True False True False
was        : True False True False
amazing    : True False False False
!          : False False False True
It         : True False True False
is         : True False True False
a          : True False True False
great      : True False False False
pleasure   : True False False False
to         : True False True False
be         : True False True False
a          : True False True False
part       : True False True False
of         : True False True False
the        : True False True False
band       : True False False False
.          : False False False True
```

In [6]:
```python
from nltk.corpus import stopwords
```

```python
In [7]: stopwords = set(stopwords.words('english'))

        stopwords
```

```
Out[7]:  {'a',
          'about',
          'above',
          'after',
          'again',
          'against',
          'ain',
          'all',
          'am',
          'an',
          'and',
          'any',
          'are',
          'aren',
          "aren't",
          'as',
          'at',
          'be',
          'because',
          'been',
          'before',
          'being',
          'below',
          'between',
          'both',
          'but',
          'by',
          'can',
          'couldn',
          "couldn't",
          'd',
          'did',
          'didn',
          "didn't",
          'do',
          'does',
          'doesn',
          "doesn't",
          'doing',
          'don',
          "don't",
          'down',
          'during',
          'each',
          'few',
          'for',
          'from',
          'further',
          'had',
          'hadn',
          "hadn't",
          'has',
          'hasn',
          "hasn't",
          'have',
          'haven',
          "haven't",
```

```
'having',
'he',
'her',
'here',
'hers',
'herself',
'him',
'himself',
'his',
'how',
'i',
'if',
'in',
'into',
'is',
'isn',
"isn't",
'it',
"it's",
'its',
'itself',
'just',
'll',
'm',
'ma',
'me',
'mightn',
"mightn't",
'more',
'most',
'mustn',
"mustn't",
'my',
'myself',
'needn',
"needn't",
'no',
'nor',
'not',
'now',
'o',
'of',
'off',
'on',
'once',
'only',
'or',
'other',
'our',
'ours',
'ourselves',
'out',
'over',
'own',
're',
's',
'same',
```

```
'shan',
"shan't",
'she',
"she's",
'should',
"should've",
'shouldn',
"shouldn't",
'so',
'some',
'such',
't',
'than',
'that',
"that'll",
'the',
'their',
'theirs',
'them',
'themselves',
'then',
'there',
'these',
'they',
'this',
'those',
'through',
'to',
'too',
'under',
'until',
'up',
've',
'very',
'was',
'wasn',
"wasn't",
'we',
'were',
'weren',
"weren't",
'what',
'when',
'where',
'which',
'while',
'who',
'whom',
'why',
'will',
'with',
'won',
"won't",
'wouldn',
"wouldn't",
'y',
'you',
```

```
            "you'd",
            "you'll",
            "you're",
            "you've",
            'your',
            'yours',
            'yourself',
            'yourselves'}
```

In [13]:
```python
test_sentence_5 = "Always do right. This will gratify some people and astonish

token_list_5 = word_tokenize(test_sentence_5)

token_list_5
```

Out[13]:
```
['Always',
 'do',
 'right',
 '.',
 'This',
 'will',
 'gratify',
 'some',
 'people',
 'and',
 'astonish',
 'the',
 'rest',
 '.']
```

In [14]:
```python
filtered_words = [w for w in token_list_5 if not w in stopwords]

filtered_words
```

Out[14]:
```
['Always', 'right', '.', 'This', 'gratify', 'people', 'astonish', 'rest',
 '.']
```

In [15]:
```python
custom_stopwords_ls = ['I', 'Harsh', 'student']
custom_stopwords = nltk.corpus.stopwords.words('english')

custom_stopwords.extend(custom_stopwords_ls)
```

In [16]:
```python
custom_stopwords
```

Out[16]:
```
['i',
 'me',
 'my',
 'myself',
 'we',
 'our',
 'ours',
 'ourselves',
 'you',
 "you're",
 "you've",
 "you'll",
 "you'd",
 'your',
 'yours',
 'yourself',
 'yourselves',
 'he',
 'him',
```

In [17]:
```python
test_sentence_6 = "The match has concluded !!! India has won the match. Will we

tokenizer = nltk.RegexpTokenizer(r"\w+")
token_list_6 = tokenizer.tokenize(test_sentence_6)
" ".join(token_list_6)
```

Out[17]: 'The match has concluded India has won the match Will we win the finals too'

In [18]:
```python
nlp = spacy.load('en_core_web_sm')
doc = nlp(test_sentence_6)
new_tokens = []
# Check if a token is a punctuation through is_punct attribute
for token in doc:
    if not token.is_punct:
        new_tokens.append(token.text)

" ".join(new_tokens)
```

Out[18]: 'The match has concluded India has won the match Will we win the finals too'

In [ ]: