Exercise - 1

In [20]:

```
!pip install nltk
```

```
Requirement already satisfied: nltk in c:\users\pavan kalyan\appdata\local\p
rograms\python\python310\lib\site-packages (3.7)
Requirement already satisfied: regex>=2021.8.3 in c:\users\pavan kalyan\appd
ata\local\programs\python\python310\lib\site-packages (from nltk) (2022.4.2
4)
Requirement already satisfied: click in c:\users\pavan kalyan\appdata\local
\programs\python\python310\lib\site-packages (from nltk) (8.1.3)
Requirement already satisfied: joblib in c:\users\pavan kalyan\appdata\local
\programs\python\python310\lib\site-packages (from nltk) (1.1.0)
Requirement already satisfied: tqdm in c:\users\pavan kalyan\appdata\local\p
rograms\python\python310\lib\site-packages (from nltk) (4.64.0)
Requirement already satisfied: colorama in c:\users\pavan kalyan\appdata\loc
al\programs\python\python310\lib\site-packages (from click->nltk) (0.4.4)
```

In [19]:

```
exp = input("Enter an Expression : ")
print(f'The Value of the Expression {exp} is',eval(exp))
```

```
Enter an Expression : 12/(4+1)
The Value of the Expression 12/(4+1) is 2.4
```

Exercise - 2(a)

In [29]:

```
my_string = "Hello, Welcome."
my_string
```

Out[29]:

```
'Hello, Welcome.'
```

In [6]:

```
print(my_string)
```

```
Hello, Welcome
```

Exercise - 2(b)

In [30]:

```
my_string + my_string
```

Out[30]:

```
'Hello, Welcome.Hello, Welcome.'
```

In [31]:

```
my_string*3
```

Out[31]:

```
'Hello, Welcome.Hello, Welcome.Hello, Welcome.'
```

In [32]:

```
my_string + " " + my_string
```

Out[32]:

```
'Hello, Welcome. Hello, Welcome.'
```

In [39]:

```
((my_string + " ")*3).strip()
```

Out[39]:

```
'Hello, Welcome. Hello, Welcome. Hello, Welcome.'
```

Exercise - 3

In [13]:

```
my_sent = ["Antartica","is","the","continent"]
s = " ".join(my_sent)
s
```

Out[13]:

```
'Antartica is the continent'
```

In [11]:

```
s.split(' ')
```

Out[11]:

```
['Antartica', 'is', 'the', 'continent']
```

Exercise - 4

In [14]:

```
import nltk
from nltk.book import text6
```

In [18]:

```
text6
```

```
<Text: Monty Python and the Holy Grail>
```

In [31]:

```
tokens = text6.tokens
list1 = list(filter(lambda x : x.endswith('ise'),tokens))
print(list1)
list2 = list(filter(lambda x : 'z' in x ,tokens))
print(list2)
list3 = list(filter(lambda x : 'pt' in x ,tokens))
print(list3)
list4 = list(filter(lambda x : x.istitle(),tokens))
print(list4[:10])
print(f'No. of Tokens : {len(tokens)}')
```

```
['wise', 'wise', 'apologise', 'surprise', 'surprise', 'surprise', 'noise',
'surprise']
['zone', 'amazes', 'Fetchez', 'Fetchez', 'zoop', 'zoo', 'zhiv', 'frozen', 'z
oosh']
['empty', 'aptly', 'Thpppppt', 'Thppt', 'Thppt', 'empty', 'Thppppt', 'temptr
ess', 'temptation', 'ptoo', 'Chapter', 'excepting', 'Thpppt']
['Whoa', 'Halt', 'Who', 'It', 'I', 'Arthur', 'Uther', 'Pendragon', 'Camelo
t', 'King']
No. of Tokens : 16967
```

Exercise - 5

In [21]:

```
sent = ['she','sells','sea','shells','by','the','sea','shore']
print("Words begin with 'sh':")
for word in sent:
    if word.startswith('sh'):
        print(word)

print("Words Contains more than 4 characters :")
for word in sent:
    if len(word)>4:
        print(word)
```

```
Words begin with 'sh':
she
shells
shore
Words Contains more than 4 characters :
sells
shells
shore
```

Exercise - 6

In [24]:

```
from nltk.book import *
sent = sent1+sent2+sent3+sent4+sent5+sent6+sent7+sent8
voc = sorted(set(sent))
print(voc[:10])
```

```
['!', ',', '-', '.', '1', '25', '29', '61', ':', 'ARTHUR']
```

Exercise - 7

In [29]:

```
import nltk
from nltk.book import text9
print(text9.index('sunset'))
print(text9[625:630])
print(text9[620:631])
```

```
629
['Park', 'lay', 'on', 'the', 'sunset']
['PARK', 'THE', 'suburb', 'of', 'Saffron', 'Park', 'lay', 'on', 'the', 'suns
et', 'side']
```

Exercise - 8

In [7]:

```
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 [11]:

```
def percent(word,text):
    lst = text.tokens
    length = len(lst)
    count = lst.count(word)
    print("Total Words in Text File : ",length)
    print(f'Count of {word} in the Text File :',count)
    per = (count/length)*100
    print(f'Percentage of Occurence of the {word} in {text} is :',round(per,3))
```

In [12]:

```
percent('the',text1)
```

```
Total Words in Text File :  260819
Count of the in the Text File : 13721
Percentage of Occurence of the the in <Text: Moby Dick by Herman Melville 18
51> is : 5.261
```

Exercise - 9

In [2]:

```
from nltk.book import text2
```

```
*** 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 [3]:

```
def vocab_size(text):
    dic = text.vocab()
    return len(dic)
```

In [4]:

```
vocab_size(text2)
```

Out[4]:

```
6833
```

Exercise - 10

In [1]:

```
import nltk
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 [12]:

```
text5
```

Out[12]:

```
<Text: Chat Corpus>
```

In [13]:

```
dic = text5.vocab()
dic
```

Out[13]:

```
FreqDist({'.': 1268, 'JOIN': 1021, 'PART': 1016, '?': 737, 'lol': 704, 'to':
658, 'i': 648, 'the': 646, 'you': 635, ',': 596, ...})
```

In [15]:

```
dic = filter(lambda x: len(x[0]) == 4,dic.items())
dic
```

Out[15]:

```
<filter at 0x17416b23a60>
```

In [16]:

```
dic = dict(dic)
sorted(dic.items(),key=lambda x:x[1],reverse=True)[:15]
```

Out[16]:

```
[('JOIN', 1021),
 ('PART', 1016),
 ('that', 274),
 ('what', 183),
 ('here', 181),
 ('....', 170),
 ('have', 164),
 ('like', 156),
 ('with', 152),
 ('chat', 142),
 ('your', 137),
 ('good', 130),
 ('just', 125),
 ('lmao', 107),
 ('know', 103)]
```

Exercise - 11

In [4]:

```
import re
from nltk.stem import PorterStemmer
from nltk import wordpunct_tokenize
ps = PorterStemmer()
```

In [15]:

```
sent = 'In the beginning God Created the heaven and the earth'
sent = re.sub(r'[^\w\s]','',sent)
tokens = wordpunct_tokenize(sent)
for token in tokens:
    print(ps.stem(token),end=" ")
```

```
in the begin god creat the heaven and the earth
```

Exercise - 12

In [9]:

```
sent = "In the beginning God created the heaven and the earth "
sent = re.sub(r'[^\w\s]','',sent)
tokens = wordpunct_tokenize(sent)
print(tokens)
print(len(tokens))
```

```
['In', 'the', 'beginning', 'God', 'created', 'the', 'heaven', 'and', 'the',
'earth']
10
```