In [2]:

```
!pip install NLTK
```

```
Requirement already satisfied: NLTK in g:\anacondaprogram\lib\site-pac
(kages (3.5
Requirement already satisfied: joblib in g:\anacondaprogram\lib\site-p
(ackages (from NLTK) (0.16.0
Requirement already satisfied: click in g:\anacondaprogram\lib\site-pa
(ckages (from NLTK) (7.1.2
Requirement already satisfied: regex in g:\anacondaprogram\lib\site-pa
(ckages (from NLTK) (2020.6.8
Requirement already satisfied: tqdm in g:\anacondaprogram\lib\site-pac
(kages (from NLTK) (4.47.0
```

In [8]:

```
import nltk
```

In [9]:

```
nltk.download('punkt')
```

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

Out[9]:

```
True
```

In [11]:

```
text=" Welcome readers. I hope you find it interesting. Please do reply."
```

In [12]:

```
from nltk.tokenize import sent_tokenize
```

In [21]:

```
print(sent_tokenize(text)) # ex 2 page 2 // how many sentecne you had  3 // how man
```

```
Welcome readers.', 'I hope you find it interesting.', 'Please do re ']
['.ply
```

In [29]:

```
import nltk
tokenizer = nltk.data.load("tokenizers/punkt/english.pickle")
```

In [30]:

```
text=" Welcome readers. I hope you find it interesting. Please do reply." # sec seq
```

In [31]:

```
tokenizer.tokenize(text) # sec seq
```

Out[31]:

```
Welcome readers.', 'I hope you find it interesting.', 'Please do re ']
['.ply
```

In [32]:

```
Arabic_text="مرحبا بكم. نحن نتعلم اساسيات مبادئ استرجاع المعلومات." # third seq
```

In [33]:

```
tokenizer.tokenize(Arabic_text) # third seq
```

Out[33]:

```
['مرحبا بكم.', 'نحن نتعلم اساسيات مبادئ استرجاع المعلومات.']
```

In [34]:

```
nltk.word_tokenize(text)
```

Out[34]:

```
,'Welcome']
,'readers'
,'.'
,'I'
,'hope'
,'you'
,'find'
,'it'
,'interesting'
,'.'
,'Please'
,'do'
,'reply'
['.'
```

In [35]:

```
tex=input() # ex 3
```

```
faisal sami alharbi
```

In [36]:

```
nltk.word_tokenize(tex)# ex 3
```

Out[36]:

```
['faisal', 'sami', 'alharbi']
```

In [37]:

```
from nltk.tokenize import TreebankWordTokenizer
```

In [38]:

```
tokenizer = TreebankWordTokenizer()
```

In [39]:

```
tokenizer.tokenize("Have a nice day. You do great!")
```

Out[39]:

```
['!' ,'Have', 'a', 'nice', 'day.', 'You', 'do', 'great']
```

In [44]:

```
from nltk.tokenize import RegexpTokenizer
```

In [48]:

```
tokenizer=RegexpTokenizer("[\w]+"+"\S+@\S+")
```

In [49]:

```
tokenizer.tokenize("Don't hesitate to ask questions or send to me your question to
```

Out[49]:

```
['faisal@gmail.com']
```

In [55]:

```
text=[" It is a pleasant evening.","Guests, who came from US arrived at the venue",
```

In [56]:

```
from nltk.tokenize import word_tokenize
```

In [57]:

```
tokenized_docs=[word_tokenize(doc) for doc in text]
```

In [58]:

```
print(tokenized_docs)
```

```
It', 'is', 'a', 'pleasant', 'evening', '.'], ['Guests', ',', 'wh']]
o', 'came', 'from', 'US', 'arrived', 'at', 'the', 'venue'], ['Food',
[['.' ,''was', 'tasty
```

In [59]:

```
import re
import string
```

In [61]:

```
x=re.compile('[%s]' % re.escape(string.punctuation))
```

In [62]:

```
tokenized_docs_no_punctuation = []
```

In [65]:

```
for review in tokenized_docs:
     new_review = []
     for token in review:
             new_token = x.sub(u'', token)
             if not new_token == u'':
                 new_review.append(new_token)
tokenized_docs_no_punctuation.append(new_review)
```

In [66]:

```
print(tokenized_docs_no_punctuation)
```

```
[['Food', 'was', 'tasty']]
```

In [67]:

```
print(text[0].upper())
```

```
.IT IS A PLEASANT EVENING
```

In [69]:

```
print(text[0].lower())
```

```
.it is a pleasant evening
```

In [5]:

```
import nltk
```

In [6]:

```
nltk.download('stopwords')
```

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

Out[6]:

```
True
```

In [7]:

```
from nltk.corpus import stopwords
```

In [8]:

```
stops=set(stopwords.words('english'))
```

In [9]:

```
words=["Don't",'hesitate','to','ask','questions']
[word for word in words if word not in stops]
```

Out[9]:

```
['Don't", 'hesitate', 'ask', 'questions"]
```

In [10]:

```
import re
from nltk.corpus import stopwords
stops=set(stopwords.words('english'))
txt= '''NLTK allows you to convert Text into
Lowercase and uppercase. Don't hesitate to ask
questions'''
wordList = re.sub("[^\w]", " ", txt).split()
[word for word in wordList if word not in stops]
```

Out[10]:

```
,'NLTK']
,'allows'
,'convert'
,'Text'
,'Lowercase'
,'uppercase'
,'Don'
,'hesitate'
,'ask'
['questions'
```

In [11]:

```
## HOMW WORK
```

In [12]:

```
import nltk
```

In [19]:

```
Faisal = open(r'C:\Users\user\Desktop\Faisal.txt')
sen=Faisal.read()
sen
```

Out[19]:

```
It is a pleasant evening.","Guests, who came from US arrived at the v'
'.enue","Food was tasty
```

In [22]:

```
import re
from nltk.corpus import stopwords
s=set(stopwords.words('english'))
wordL=re.sub("[^\w]"," ",sen).split()
[word for word in wordL if word not in stops]
```

Out[22]:

```
            ,'It']
            ,'pleasant'
            ,'evening'
            ,'Guests'
            ,'came'
            ,'US'
            ,'arrived'
            ,'venue'
            ,'Food'
            ['tasty'
```

In [ ]:

```

```