:[21] In

```
! pip install nltk
```

```
Requirement already satisfied: nltk in d:\users\d7me_\anaconda3\lib\site-pac
(kages (3.4.5
Requirement already satisfied: six in d:\users\d7me_\anaconda3\lib\site-pack
(ages (from nltk) (1.12.0
```

:[22] In

```
import nltk as ntk
```

:[23] In

```
ntk.download('punkt')
```

```
nltk_data] Downloading package punkt to]
...nltk_data]     C:\Users\D7me_\AppData\Roaming\nltk_data]
!nltk_data]   Package punkt is already up-to-date]
```

Out[23]:

```
True
```

:[24] In

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

:[25] In

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

:[26] In

```
print(sent_tokenize(text))
```

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

:[27] In

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

:[28] In

```
tokenizer.tokenize(text)
```

Out[28]:

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

In [29]:

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

Out[29]:

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

In [31]:

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

In [32]:

```
tokenizer.tokenize(Arabic_text)
```

Out[32]:

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

In [2]:

```
Arabic=input("Please write a text")
```

```
اهلا وسهلاPlease write a text
```

In [3]:

```
import nltk
from nltk.tokenize import TreebankWordTokenizer
tokenizer = TreebankWordTokenizer()
tokenizer.tokenize("Have a nice day. You do great!")
```

Out[3]:

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

:[4] In

```
from nltk.tokenize import RegexpTokenizer
tokenizer=RegexpTokenizer("[\w]+")
tokenizer.tokenize('''Don't hesitate to ask
questions or send to me your question to
mohsarem@gmail.com''')
```

Out[4]:

```
,'Don']
,'t'
,'hesitate'
,'to'
,'ask'
,'questions'
,'or'
,'send'
,'to'
,'me'
,'your'
,'question'
,'to'
,'mohsarem'
,'gmail'
['com'
```

:[5] In

```
tokenizer=RegexpTokenizer("\S+@\S+")
tokenizer.tokenize('''Don't hesitate to ask
questions or send to me your question to
mohsarem@gmail.com''')
```

Out[5]:

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

:[6] In

```
text=[" It is a pleasant evening.","Guests, who came from US arrived at the venue","Food wa
from nltk.tokenize import word_tokenize
tokenized_docs=[word_tokenize(doc) for doc in text]
print(tokenized_docs)
```

```
It', 'is', 'a', 'pleasant', 'evening', '.'], ['Guests', ',', 'who', 'cam']]
 ,'e', 'from', 'US', 'arrived', 'at', 'the', 'venue'], ['Food', 'was', 'tasty
[['.'
```

:[7] In

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

print(tokenized_docs_no_punctuation)
```

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

:[8] In

```
text= "NLTK allows you to convert Text into Lowercase and uppercase"
print(text.upper())
print(text.lower())
```

```
NLTK ALLOWS YOU TO CONVERT TEXT INTO LOWERCASE AND UPPERCASE
nltk allows you to convert text into lowercase and uppercase
```

:[9] In

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

```
nltk_data] Downloading package stopwords to]
...nltk_data]     C:\Users\D7me_\AppData\Roaming\nltk_data]
.nltk_data]   Unzipping corpora\stopwords.zip]
```

Out[9]:

```
True
```

:[10] In

```
import nltk
from nltk.corpus import stopwords
stops=set(stopwords.words('english'))
words=["Don't",'hesitate','to','ask','questions']
[word for word in words if word not in stops]
```

Out[10]:

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

:[11] In

```
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[11]:

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

:[18] In

```
import nltk
Abdulrhman = open(r'D:\Users\D7me_\Anaconda3\Abdulrhman.txt')
sen = Abdulrhman.read()
sen
```

Out[18]:

```
In computer science, artificial intelligence (AI), sometimes called mach'
 ine intelligence, is intelligence demonstrated by machines, incontrast to
the natural intelligence displayed by humans and animals. Computer scienc
 e defines AI research as the study of intelligent agents: any device that
perceives its environment and takes actions that maximize its chance of s
'uccessfullyachieving its goals
```

HW2

:[19] In

```
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[19]:

```
             ,'In']
             ,'computer'
             ,'science'
             ,'artificial'
             ,'intelligence'
             ,'AI'
             ,'sometimes'
             ,'called'
             ,'machine'
             ,'intelligence'
             ,'intelligence'
             ,'demonstrated'
             ,'machines'
             ,'incontrast'
             ,'natural'
             ,'intelligence'
             ,'displayed'
             ,'humans'
             ,'animals'
             ,'Computer'
             ,'science'
             ,'defines'
             ,'AI'
             ,'research'
             ,'study'
             ,'intelligent'
             ,'agents'
             ,'device'
             ,'perceives'
             ,'environment'
             ,'takes'
             ,'actions'
             ,'maximize'
             ,'chance'
             ,'successfullyachieving'
             ['goals'
```

:[ ] In

```

```