In [2]:
```
import re # regular expression library
import nltk # natural language toolkit library
import string
```

```
  File "C:\Users\MEHDIH~1\AppData\Local\Temp/ipykernel_5108/227190490.py", line 4
    jupyter nbconvert --to webpdf --allow-chromium-download Untitled.ipynb
            ^
SyntaxError: invalid syntax
```

In [27]:
```
original_text = """Artificial intelligence is human like intelligence.
                   It is the study of intelligent artificial agents.
                   Science and engineering to produce intelligent machines.
                   Solve problems and have intelligence.
                   Related to intelligent behavior.
                   Developing of reasoning machines.
                   Learn from mistakes and successes.
                   Artificial Intelligence is related to reasoning in everyday situations.
```

In [28]:
```
original_text
```

Out[28]:
```
'Artificial intelligence is human like intelligence.\n                   It is the study o
f intelligent artificial agents.\n                   Science and engineering to produce in
telligent machines.\n                   Solve problems and have intelligence.\n
        Related to intelligent behavior.\n                   Developing of reasoning machi
nes.\n                   Learn from mistakes and successes.\n                   Artificial
Intelligence is related to reasoning in everyday situations.'
```

In [29]:
```
original_text = re.sub(r'\s+', ' ', original_text)
```

In [46]:
```
original_text
```

Out[46]:
```
'Artificial intelligence is human like intelligence. It is the study of intelligent artifi
cial agents. Science and engineering to produce intelligent machines. Solve problems and h
ave intelligence. Related to intelligent behavior. Developing of reasoning machines. Learn
from mistakes and successes. Artificial Intelligence is related to reasoning in everyday s
ituations.'
```

In [47]:
```
nltk.download('punkt')
```

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

Out[47]:
```
True
```

In [56]:
```
nltk.download('stopwords')
```

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

Out[56]:
```
True
```

In [76]:
```
stopwords = nltk.corpus.stopwords.words('english') #these are the words in english that ha
#print(stopwords)
#len(stopwords)
```

In [78]:
```
string.punctuation
```

Out[78]:
```
'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

In [83]:
```
def preprocess(text): # defining our function and it recieves Text as parameter and will c
    formatted_text = text.lower() # this function lower here makes the texts in Lowercase

    tokens = [] # made new variable which includes each one of the words we have in the su
    for token in nltk.word_tokenize(formatted_text):
        tokens.append(token)
        #print(tokens)

    tokens = [word for word in tokens if word not in stopwords and word not in string.punc


    formatted_text = ' '.join(element for element in tokens)

    return formatted_text
```

In [84]:
```
formatted_text = preprocess(original_text) #now make a variable to implement our function
formatted_text
```

Out[84]:
```
'artificial intelligence human like intelligence study intelligent artificial agents scien
ce engineering produce intelligent machines solve problems intelligence related intelligen
t behavior developing reasoning machines learn mistakes successes artificial intelligence
related reasoning everyday situations'
```

In [ ]:

In [ ]: