```
pip install nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk) (8.1.8)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk) (4.67.1)
```

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

```
import nltk
nltk.download('punkt_tab')
```

```
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Package punkt_tab is already up-to-date!
True
```

```
text =  "Hello There!, myself :- Mudit Garg , Hello how are you?"
```

```
sent_tokenize(text)
```

```
['Hello There!, myself :- Mudit Garg , Hello how are you?']
```

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

```
text1 =  "I am 20 years old. I am currently studing in Manav Rachna"
```

```
word_tokenize(text1)
```

```
['I',
 'am',
 '20',
 'years',
 'old',
 '.',
 'I',
 'am',
 'currently',
 'studing',
 'in',
 'Manav',
 'Rachna']
```

```
from nltk.corpus import stopwords
nltk.download('stopwords')
print(stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourse
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```
text.upper()
```

```
lower = text.lower()
lower
```

```
text.split()
```

```
['Hello',
 'There!,',
 'myself',
 ':-',
 'Mudit',
 'Garg',
 ',',
 'Hello',
 'how',
 'are',
 'you?']
```

```python
new_text = text.replace("Hello","hi")
new_text
```

```python
count = text.count("e")
count
```

> 6

STRINGS

```python
length=len(text)
length
```

> 55

```python
add=text + text1
add
```

```python
repeat_text=text*3
repeat_text
```

```python
first_char=text[4]
first_char
```

```python
Slicing=text[:30]
Slicing
```

```python
Slicing2=text[::2]
Slicing2
```

```python
Slicing2=text[::-2]
Slicing2
```

> '?o r o le rGtdM- lsm,eeTolH'

```python
Slicing3=text[:18:2]
Slicing3
```

> 'HloTee,ms'

```python
Slicing4=text[:27:-1]
Slicing4
```

> '?uoy era woh olleH , graG t'

REMOVING STOPWORDS

```python
stop_words = set(stopwords.words('english'))
words = word_tokenize(lower)
after_remove_stopwords = [word for word in words if word.lower() not in stop_words]
print(lower)
print("after removing stopwords:", after_remove_stopwords)
```

> hello there!, myself :- mudit garg , hello how are you?
> after removing stopwords: ['hello', '!', ',', ':', '-', 'mudit', 'garg', ',', 'hello', '?']

```python
text = "Myself Mudit Garg, Currently i am pursuing a bachelor degree from in Computer Science and Engineering from Manav Rachna Universit

words = word_tokenize(text)

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

```python
filtered_words = [word for word in words if word.lower() not in stop_words]

filtered_text = ' '.join(filtered_words)

print("Original Text:", text)
print("Filtered Text:", filtered_text)
```

⇥ Original Text: Myself Mudit Garg, Currently i am pursuing a bachelor degree from in Computer Science and Engineering from Manav Racl
   Filtered Text: Mudit Garg , Currently pursuing bachelor degree Computer Science Engineering Manav Rachna University

◀ ━━━ ▶

```python
def find_top_10_words(text):
    words = word_tokenize(text)

    stop_words = set(stopwords.words('english'))

    filtered_words = [word.lower() for word in words if word.isalpha() and word.lower() not in stop_words]

    word_freq = Counter(filtered_words)

    top_10 = word_freq.most_common(10)

    return top_10

text = """
Myself Mudit Garg, Currently i am pursuing a bachelor degree from in Computer Science and Engineering from Manav Rachna University
"""

result = find_top_10_words(text)
print("Top 10 Words (Excluding Stopwords):", result)
```

⇥ Top 10 Words (Excluding Stopwords): [('mudit', 1), ('garg', 1), ('currently', 1), ('pursuing', 1), ('bachelor', 1), ('degree', 1),

◀ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ▶