

```
#EXCLLER DAY-24(198)
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)
```

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

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

```
sample_text.encode("utf-8")
```

```
b'\nOnce upon a time in a small village, there lived a young girl named Alice. \nShe loved exploring the world around her and
discovering new things.\n'
```

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

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

```
import nltk
from nltk.tokenize import word_tokenize
from nltk.probability import FreqDist
```

```
# Ensure NLTK data is downloaded
nltk.download('punkt')
```

```
# Define sample text
sample_text = """
Once upon a time in a small village, there lived a young girl named Alice.
She loved exploring the world around her and discovering new things.
"""
```

```
# Tokenize the text
try:
    tokens = word_tokenize(sample_text) # ✅ Fix potential tokenization issues
except Exception as e:
    print(f"Error tokenizing text: {e}")
    tokens = []
```

```
# Compute word frequency distribution only if tokenization is successful
if tokens:
    freq_dist = FreqDist(tokens)
```

```
# Get the 10 most common words
most_common_words = freq_dist.most_common(10)
```

```
# Print the most common words and their frequencies
print("Most Common Words and Their Frequencies:")
for word, frequency in most_common_words:
    print(f'{word}: {frequency}')
```

```
else:
    print("No tokens generated, check input text.")
```

```
Error tokenizing text:
*****
Resource punkt_tab not found.
Please use the NLTK Downloader to obtain the resource:

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

For more information see: https://www.nltk.org/data.html

Attempted to load tokenizers/punkt_tab/english/

Searched in:
- '/root/nltk_data'
```

```
- '/usr/nltk_data'  
- '/usr/share/nltk_data'  
- '/usr/lib/nltk_data'  
- '/usr/share/nltk_data'  
- '/usr/local/share/nltk_data'  
- '/usr/lib/nltk_data'  
- '/usr/local/lib/nltk_data'
```

No tokens generated, check input text.

[nltk_data] Downloading package punkt to /root/nltk_data...

[nltk_data] Package punkt is already up-to-date!