

SALONI BHINGARDIVE ROLL NO. 23 CSE (DS) NLP PRACTICAL 01

```
import nltk
nltk.download('punkt')
from nltk.tokenize import word_tokenize
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
```

```
text= "I'm currently reading THE CONCLAVE by Robert Harris. Its's a pretty detailed and thrilling description of what takes place after the H
```

```
words=word_tokenize(text)
print(words)
```

```
['I', 'm', 'currently', 'reading', 'THE', 'CONCLAVE', 'by', 'Robert', 'Harris', '.', 'Its', 's', 'a', 'pretty', 'detailed', 'and', 'th
```

```
from nltk.tokenize import sent_tokenize
sentence=sent_tokenize(text)
print(sentence)
```

```
["I'm currently reading THE CONCLAVE by Robert Harris.", "Its's a pretty detailed and thrilling description of what takes place after th
```

```
sentence[0]
```

```
'I'm currently reading THE CONCLAVE by Robert Harris.'
```

```
print(words[8])
```

```
Harris
```

```
from nltk.tokenize import sent_tokenize, word_tokenize
```

```
text2="Harry lay flat on his back, breathing hard as though he had been running. He had awoken from a vivid dream with his hands pressed ov
for_words=word_tokenize(text2)
for_sentence=sent_tokenize(text2)
```

```
print(for_words)
print(for_sentence)
```

```
['Harry', 'lay', 'flat', 'on', 'his', 'back', ',', 'breathing', 'hard', 'as', 'though', 'he', 'had', 'been', 'running', '.', 'He', 'had'
['Harry lay flat on his back, breathing hard as though he had been running.', 'He had awoken from a vivid dream with his hands pressed c
```

```
print(for_words[0])
```

```
Harry
```

```
print(for_sentence[7])
```

```
He turned on the lamp beside him, scrambled out of bed, crossed the room, opened his wardrobe, and peered into the mirror on the inside
```

```
text3= "Saloni's habit of reading is a good one."
for_words_new=word_tokenize(text3)
print(for_words_new)
#s is taken as a different token
```

```
['Saloni', 's', 'habit', 'of', 'reading', 'is', 'a', 'good', 'one', '.']
```

```
text4= "Saloni_Bhingardive is a student."
for_words_new2=word_tokenize(text4)
print(for_words_new2)
#a_b is is taken as one single token
#full stop is taken as a separate token
```

```
↳ ['Saloni_Bhingardive', 'is', 'a', 'student', '.']
```

Start coding or [generate](#) with AI.

Start coding or [generate](#) with AI.