[Day-8 2211cs020208] Write a python script that:1.Tokenizes a sample paragraph into words and sentence

In [*]:
```
!pip install nltk
import nltk
nltk.download('punkt')
def tokenize_text(paragraph):
    sentences = nltk.sent_tokenize(paragraph)
    words = [nltk.word_tokenize(sentence) for sentence in sentences]
    return sentences, words
paragraph = """
Tokenization is the process of breaking text into smaller units called to
These tokens can be words, sentences, or even smaller units. Tokenization
important step in text preprocessing.
"""
sentences, words = tokenize_text(paragraph)
print("Sentences:")
for sentence in sentences:
    print(sentence)
print("\nWords:")
for word_list in words:
    print(word_list)
```

```
Requirement already satisfied: nltk in c:\users\nihar\anaconda3\lib\site-pac
kages (3.7)
Requirement already satisfied: click in c:\users\nihar\anaconda3\lib\site-pa
ckages (from nltk) (8.0.4)
Requirement already satisfied: joblib in c:\users\nihar\anaconda3\lib\site-p
ackages (from nltk) (1.1.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\nihar\anaconda3\l
ib\site-packages (from nltk) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\nihar\anaconda3\lib\site-pac
kages (from nltk) (4.64.1)
Requirement already satisfied: colorama in c:\users\nihar\anaconda3\lib\site
-packages (from click->nltk) (0.4.6)
```

In [ ]:
```
```