Aim: Apply Various text preprocessing techniques tokenization and stop word removal.

Objective: To create sentence and word tokens from given paragraph.

Theory:
Tokenization is process of tokenizing or splitting a string, text into token. One can think of token as parts like word is a token in sentence and sentence is token in paragraph.

Program -

```
from nltk import word_tokenize

from nltk import sent_tokenize

from nltk import wordpunct_tokenize

from nltk.corpus import stopwords

t = "The sun is shining. I go out for a walk."

p = "One dollar and eighty-seven cents. That was all. And sixty cents of it was in pennies. Pennies saved one and two at a time by bulldozing the grocer and the vegetable man and the butcher until one's cheeks burned with the silent imputation of parsimony that such close dealing implied. One dollar and eighty-seven cents. And the next day would be Christmas..."

wt = word_tokenize(t)

print(wordpunct_tokenize(t))

print(sent_tokenize(p))

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

f=[]

for i in wt:

        i = i.lower()

        if i not in stopwords:

                f.append(i)

print(f)
```

Output - ['The', 'sun', 'is', 'shining', '.', 'I', 'go', 'out', 'for', 'a', 'walk', '.']

['One dollar and eighty-seven cents.', 'That was all.', 'And sixty cents of it was in pennies.', 'Pennies saved one and two at a time by bulldozing the grocer and the vegetable man and the butcher until one's cheeks burned with the silent imputation of parsimony that such close dealing implied.', 'One dollar and eighty-seven cents.', 'And the next day would be Christmas...']

['sun', 'shining', '.', 'go', 'walk', '.']

Conclusion: Tokenization is a critical step in natural language processing (NLP). It involves breaking text into smaller units or tokens like words or subwords. This process is vital for NLP tasks like machine translation and sentiment analysis. Accurate tokenization is crucial for language models and NLP algorithms to work effectively, making it a cornerstone of modern NLP and enabling applications reliant on textual data.