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: Natural language processing (NLP)'s essential and crucial stage of tokenization is one that cannot be overlooked. It entails segmenting text into more manageable pieces or tokens, like words or subwords. This procedure not only makes it possible for computers to comprehend and analyse human language, but it also plays a significant part in a number of NLP activities like text classification, sentiment analysis, and machine translation. Accurate tokenization is essential to modern NLP because it ensures that language models and algorithms can function properly and makes it easier to create applications that depend on textual input.