Experiment 3

Aim:

To tokenize a text file into sentences and words and remove stop words from a file.

Theory:

Tokenization involves breaking the text file into meaningful units to facilitate analysis.

Stop words are commonly occurring words that often carry little meaning and can be often excluded from text analysis to improve its quality.

The resulting processed files contains meaningful sentences and words, free from stop words and is ready for further analysis.

Flowchart: