

AIM: To parse the dataset consisting of words and to display the 10 most frequent words

CODE:

```
% Import the data file
text_file = fileread("Example_text.txt");

% Convert 'text_file' to a string
text_file = string(text_file);

% remove unwanted characters
% specify which characters to remove
characters = {'.', ':', '/', '"', ','};
% replace those characters from the text file with empty space
text_file = replace(text_file, characters, ' ');

% use split function to split each string into individual words
text_file = split(text_file);

% use tabulate function to get frequency of each word in a tabular format
frequencies = tabulate(text_file);

% arrange in descending order
[ freq, index] = sort(cell2mat(frequencies(:,3)), 'descend');

% create a barplot of 10 most frequently occurring words
% x-axis should have names of words
% y-axis should have percentage frequencies of the words
% x-axis requires categories. since tabulate returns an array, categorical
% function converts it to a category.
% y-axis requires numeric. cell2mat converts to numeric
%
bar(categorical(frequencies(index(1:10),1)),
cell2mat(frequencies(index(1:10), 3)));
```

FIG. 1:

