ASSIGNMENT 2

Word Frequency Counter

Write a program that reads a text file and counts the number of times each word appears(frequency). Your program should then output the words in decreasing order of their frequency. The user should also be able to specify a threshold, and your program should output only those words whose frequency is more than the threshold.

Your program should run in O(n+klogk), where n is the total no. of words in text file and k is the no. of distinct words having frequency more than the threshold.

To do this, you will need to design/implement an efficient hashing and sorting algorithm. You can use tricks like probing, chaining etc to make the hashing efficient. Do not use the built in Java classes for hashing or sorting (ie. HashMap, HashSet, etc). Your program should take the input from a file test.txt and prompt the user to enter the threshold. The output should be printed on the terminal.

You should consolidate words by removing capitalization and split words by spaces and punctuation. Remove any punctuation that splits words, but keep the apostrophe since it often signifies contractions.

Example:

Input:

Text : "The function of a paragraph is to mark a pause, setting the paragraph apart from what precedes it. If a paragraph is preceded by a title or subhead, the indent is superfluous and can therefore be omitted."

Threshold : 2

Output:

```
a  :  4
is :  3
paragraph: 3
the : 3
```