CSE 225 - Data Structures and Algorithms
Assignment 1
Due Date: 12-07-2017

Total marks: 50

Submission Instructions:

Email your code to nabeel.mohammed.for.nsu@gmail.com before 11:59pm of 12-07-2017. No extensions will be given. Late submissions will be penalised at a rate of 10 marks per day late.

Task:

For this assignment you are required to write a program which accepts the name of a text file from the command line. You can assume that the text file contains English text. The program should then output the frequency of each word in the file, along with a list of position indexes. The information for each unique word should be shown only once.

If for example the input text file contains the following content:

> The quick brown fox jumps over the lazy dog. However, the dog really did not care.

The output should be the following:

```
Total number of words: 16

Word frequencies and positions:

the − 0.1875: 1,7,11
quick − 0.0625: 2
brown − 0.0625: 3
fox − 0.0625: 4
jumps − 0.0625: 5
over − 0.0625: 6
lazy − 0.0625: 8
dog − 0.125: 9, 12
however − 0.0625: 10
really − 0.0625: 13
did − 0.0625: 14
not − 0.0625: 15
care − 0.0625: 16
```

Please note that this example is extremely small. Your program should be written so that it can handle an arbitrary number of words, appearing in an arbitrary number of positions. You can also assume that to receive good marks, your program must reliably work with files larger than 1 MB.