AI PRAC: Natural Language Processing

Many text parsing operations in Natural Language Processing are statistically based, often as a precursor to further processing. This practical requires you to enhance some given code so that it returns various statistical properties of text.

Download the file **TextAnalyser.cs** and **Treasureisland.txt** from Canvas. Set up your project and ensure the text file is in the project's Debug folder. On compiling and running the program the contents of the text file are read and stored in string variable *input* and then *input* is displayed to screen (suppress this output with a '//' at the appropriate line in code if you wish).

1. Develop **five methods** that must be called from Main() to analyze the text. In each case *input* is passed to the respective method. The five methods must in each case return>

    a) the number of printable characters (including apostrophes, question marks, etc)
    b) the number of white space (or non-printing) characters
    c) the number of vowels
    d) the number of consonants
    e) the frequency of each vowel

    Methods a)–d) should return a single statistical count.
    Method e) *should not actually need to 'return' anything ie make it void*. Instead create an empty array of 5 ints in Main() each initialised to 0 and pass the array to the method along with *input*. The method should then update the array elements – on control returning to Main() the modified contents should be accessible from Main() due to the way in which C# array references (or indeed references to any C# objects) are passed to methods.

2. Print the results of each method call to screen from Main().

You will probably need to web-search some operations, including i) array declarations and loop processing in C# ii) how to derive lowercase strings from mixed-case strings iii) how to identify whitespace characters. Often the character recognition/comparison is done most efficiently by making use of consecutive sequences of integer ASCII codes rather than the character symbols themselves. A listing of ASCII codes is given on Canvas for use with this practical.