CSCI 2020 - ASSIGNMENT 2 REPORT

February 22nd, 2015

Luisa Rojas Garcia

The assignment consisted in coding two readers: buffered and unbuffered. The readers were supposed to get the last letter of every word in the dictionary. The number of times each letter from A-Z appeared in said spot was to be recorded and then made into a histogram. Once the coding was done, the efficiency of the buffered program was tested on a hard disk drive (HDD) and on a USB.

In order to achieve this, a couple more methods than expected needed to be added. In the case of the unbuffered reader, the following methods were used: run(), printHistogram(int[] h) and getLastCharOfLine(Reader r). The first method was the one being called by the main function, it used the other two methods. On the other hand, the buffered reader was a lot easier to code and only needed two methods: run() and printHistogram(). The results were written on a text file and then converted into an excel file to plot them.

The time it took these two readers to run the program was compared and it was noted that the buffered reader is a lot more optimal but only with large buffer sizes; however, it is important to mention that buffer sizes that are too large may also harm the performance of the program since it would occupy more memory than the necessary.

Once the both readers were outputting the correct values in the form of a histogram, they were both tested on a HDD and on a USB 1000 times, increasing the buffer size by 1 each time. The expected result was for the time it took the HDD to run the program to be significantly lower than the time it took the USB to run the same file. Surprisingly, both methods resulted in very similar outputs – refer to the graphs below.

The test was ran 2 times to lower the chances of there being any errors. There are some visible peaks both times on both the HDD and the USB. The cause of these peaks is not yet clear, but I assume it must have to be something related to the usage of the computer at the time the program was ran (said usage was purposely minimal in both cases).