For my attempt at creating a Burrows-Wheeler Transform tool I set up 1 main class and 2 classes to actually house the methods for doing the transform in either direction. One class is for the Forward Transform and the other class does the Reverse. Both Classes read and write their respective input and output files. Writing is done on a segment by segment basis; i.e. If a value of 100 is passed to the BWT it will cut out 100 char strips from the original text and perform the transform on them. In reverse a similar principle applies, however, as I saved the Array location for the original string in-line with the BWT strings, Instead of using the length passed to it, the reverse uses the location of the > character to determine the end of the segment. As < and > did not appear in the original text, I utilized them to house the Array location for the untransformed string. My longest run counter was borrowed from stackoverflow.com and the source is provided in the code.

The program is implemented in full as prescribed in the instructions. All text was returned to its original form after BWT was run. It does not however return to normal after compression as somewhere in either MTF or Huffman there is a language change.

Run Outputs:

V:\proj\_src> java BWT + 20 MobyDick.txt MDBWT20.txt

BWT Runtime: 0.5000s

Longest run: 14, for the character " "

V:\proj\_src> java BWT + 80 MobyDick.txt MDBWT80.txt

BWT Runtime: 1.1410s

Longest run: 21, for the character " "

V:\proj\_src> java BWT + 500 MobyDick.txt MDBWT500.txt

BWT Runtime: 3.1680s

Longest run: 74, for the character " "

V:\proj\_src> java BWT - 20 MDBWT20.txt Undo20.txt

Reverse BWT Runtime: 2.4840s

V:\proj\_src> java BWT - 80 MDBWT80.txt Undo80.txt

Reverse BWT Runtime: 16.6160s

V:\proj\_src> java BWT - 500 MDBWT500.txt Undo500.txt

Reverse BWT Runtime: 246.9790s

V:\proj\_src>

|  |  |
| --- | --- |
| File | Size |
| MobyDick.txt | 1228 KB |
| MDBWT20 | 1443 KB |
| MDBWT80 | 1288 KB |
| MDBWT500.txt | 1240 KB |
| MDHuff20.comp | 959 KB |
| MDHuff80.comp | 824 KB |
| MDHuff500.comp | 696 KB |
| MobyDick.7z | 408 KB |

Compression Ratios:

Original: 1228 KB

Huffman with MTF & BWT 20:

Huffman with MTF & BWT 80:

Huffman with MTF & BWT 500:

7-Zip Compression: