Empirical Analysis

Table 1 and Table 2 show the HuffMark analysis for compressing text files and binary files, respectively. Using BurrowsWheeler, the files can be compressed into much smaller files comparing with the original SimpleHuff compression, although the time lengths for compression increase. The BurrowsWheeler is much more effective in compression of the binary files.

Table 1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Text files in Calgary | from | Original SimpleHuffProcessor | | | BurrowsWheeler  SimpleHuffProcessor | | |
| to | time | to/from | to | time | to/from |
| bib | 111261 | 73791 | 0.215 | 0.66 | 42467 | 0.533 | 0.38 |
| book1 | 768771 | 439405 | 0.89 | 0.57 | 343299 | 2.376 | 0.45 |
| book2 | 610856 | 369331 | 0.688 | 0.60 | 240476 | 0.9 | 0.39 |
| geo | 102400 | 73588 | 0.194 | 0.72 | 74752 | 0.242 | 0.73 |
| news | 377109 | 247424 | 0.483 | 0.66 | 169595 | 0.587 | 0.45 |
| obj1 | 21504 | 17081 | 0.08 | 0.79 | 12475 | 0.107 | 0.58 |
| obj2 | 246814 | 195127 | 0.366 | 0.79 | 99893 | 0.369 | 0.40 |
| paper1 | 53161 | 34367 | 0.113 | 0.65 | 21228 | 0.119 | 0.40 |
| paper2 | 82199 | 48645 | 0.093 | 0.59 | 33492 | 0.122 | 0.41 |
| paper3 | 46526 | 28305 | 0.104 | 0.61 | 20034 | 0.07 | 0.43 |
| paper4 | 13286 | 8890 | 0.017 | 0.67 | 6364 | 0.019 | 0.48 |
| paper5 | 11954 | 8461 | 0.018 | 0.71 | 5931 | 0.018 | 0.50 |
| paper6 | 38105 | 25053 | 0.057 | 0.66 | 15564 | 0.054 | 0.41 |
| pic | 513216 | 107582 | 0.233 | 0.21 | 116451 | 13.759 | 0.23 |
| progc | 39611 | 26944 | 0.097 | 0.68 | 16025 | 0.101 | 0.40 |
| progl | 71646 | 44013 | 0.107 | 0.61 | 22430 | 0.112 | 0.31 |
| progp | 49379 | 31244 | 0.09 | 0.63 | 15300 | 0.084 | 0.31 |
| trans | 93695 | 66248 | 0.127 | 0.71 | 28960 | 0.153 | 0.31 |
| total bytes read: | | 3251493 | | | 3251493 | | |
| total compressed bytes | | 1845499 | | | 1282304 | | |
| total percent compression | | 43.241 | | | 60.563 | | |
| compression time: | | 3.972 | | | 19.725 | | |

Table 2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| binary files in Waterloo | from | Original SimpleHuffProcessor | | | BurrowsWheeler  SimpleHuffProcessor | | |
| to | time | to/from | to | time | to/from |
| clegg.tif | 2149096 | 2034591 | 3.839 | 0.95 | 967115 | 21.109 | 0.45 |
| frymire.tif | 3706306 | 2188589 | 4.37 | 0.59 | 720455 | 35.899 | 0.19 |
| lena.tif | 786568 | 766142 | 1.429 | 0.97 | 665511 | 1.941 | 0.85 |
| monarch.tif | 1179784 | 1109969 | 2.076 | 0.94 | 862550 | 2.357 | 0.73 |
| peppers.tif | 786568 | 756964 | 1.443 | 0.96 | 652441 | 1.843 | 0.83 |
| sail.tif | 1179784 | 1085497 | 2.072 | 0.92 | 931162 | 2.436 | 0.79 |
| serrano.tif | 1498414 | 1127641 | 2.162 | 0.75 | 286581 | 4.634 | 0.19 |
| tulips.tif | 1179784 | 1135857 | 2.102 | 0.96 | 998964 | 2.636 | 0.85 |
| total bytes read: | | 12466304 | | | 12466304 | | |
| total compressed bytes | | 10205250 | | | 6082347 | | |
| total percent compression | | 18.137 | | | 51.210 | | |
| compression time: | | 19.493 | | | 72.855 | | |