sievenna Testing

Onni Aarne

January 21, 2018

1 Unit Testing

A link to an up-to-date report can be found in README.md.

2 Performance Testing

sievenna's Huffman coding has been tested on the Large Text Compression Benchmark [2] as well as an uncompressed tarball of the standard Calgary corpus [1].

Image compression benchmark used was an 8-bit RGB photograph called nightshot_iso_100 from imagecompression.info [3].

Results achieved can be seen in table 1. The software fares best with text, achieving a compression ratio of about 1.5. The different sized variants of the Large Text Compression benchmark combined with the relatively small Calgary corpus show that the ratios achieved are not heavily dependent on input size.

If the reader wishes to replicate these results, among the project's test HuffmanCoderTest and CalgaryTest test the performace of the program on the nightshot and the Calgary corpus respectively. The Large Text Compression benchmark is too large to include in the repository, but can be found here.

File	Size	Compressed	Ratio	Comp. Time	Decomp. Time
nightshot_iso_100.ppm	22.128 MB	17.843427 MB	1.240	2.654 s	1.989 s
enwik8	100 MB	63.862	1.566	9.100 s	6.900 s
enwik9	1000 MB	648.370 MB	1.542	68 s	59 s
Calgary Corpus	3.154 MB	2.125 MB	1.484	0.385 s	0.321 s

Table 1: Performance statistics for Huffman coding. Note: enwik8 and enwik9 are different sizes of the Large Text Compression Benchmark.

References

[1] Timothy Bell, Ian H Witten, and John G Cleary. Modeling for text compression. **ACM Computing Surveys (CSUR)**, 21(4):557–591, 1989.

[2] Matt Mahoney. Large text compression benchmark. `http://mattmahoney.net/dc/text.html`, 2011.

[3] Rawzor. The new test images. `http://imagecompression.info/test_images/`, 2008. Accessed 26.12.2017.