In this paper, we study different methods of data compression algorithms on English text files: LZW, Huffman, Fixed-length code (FLC), and Huffman after using Fixed-length code (HFLC). We evaluate and test these algorithms on different text files of different sizes and make a comparison in terms of compression: Size, Ratio, Time (Speed), and Entropy. We found that
LZW is the best algorithm in all of the compression scales that we tested, then Huffman, Huffman after using Fixed-length code (HFLC), and Fixed-length code (FLC), respectively. The Entropy for them was: 4.719, 4.855, 5.014, and 6.889 respectively, for the sample tested files.

Reference

- Debra A. Lelewer and Daniel S. Hirschberg, Data Compression www.ics.uci.edu/~dan/pubs/DataCompression.html
- 7 Ellen Chang, Udara Fernando, and Jane Hu, Data Compression http://www.stanford.edu/~udara/SOCO/lossless/index.htm
- Mark Nelson, LZW Data Compression, Dr. Dobb's Journal, October 1989 www.dogma.net/markn/articles/lzw/lzw.htm
- Matt Powell, University of Canterbury, last updated November 20, 2001 http://corpus.canterbury.ac.nz
- Mitsuharu ARIMURA, Bookmarks on Source Coding/Data Compression, 2001 http://www.hn.is.uec.ac.jp/~arimura/compression_links.html
- Owen L. Astrachan, Huffman Coding : ACS2 Assignment From ASCII Coding to Huffman Coding, Feb 2004 www.cs.duke.edu/csed/poop/huff/info/
- Steve Linton, Data Compression, Information Sources. www.dcs.st-and.ac.uk/~sal/school/cs3010/lectures/forhtml/node2.html
- Tore Nestenius, Huffman Trees for Data Compression, 2004. www.programmersheaven.com/2/Art_Huffman_p1

Index Terms

Computer Science Data Structures

Key words

Data Compression Source Mapping Huffman

Coding Hamming

Entropy

LZW