Lempel-Ziv Compression

Wenyu Liang

March 13, 2023

1 Program description

This program will implement Lempel-Ziv Compression algorithm, LZ78 specifically. This is a dictionary-based algorithm, where the key is a prefix, also known as a word, and the value is a code that we can utilize for fast look-ups, during compression. The decompression is very similar, except that the key and value for the decompressing dictionary is swapped. Users can compress and decompress their data with this program.

2 What I learned

1. **Algorithm:** I learned how the LZ78 compression work.
2. **IO:** I learned how to write bit by bit into a byte array. Plus, I learned the concept endianness. I learned to use flush to write out the buffer that's not filled.
3. **trie:** I learned a new data structure trie which is useful to look for strings that may have identical prefix.

3 Relation between entropy and compression efficiency

If we look at this small dataset that I test, we could find that the compression ration decreases as the entropy of the original file increases.

File name	Entropy	Compression ratio
file1	4.950854	62.81%
file2	4.526927	38.48%
file3	4.255492	17.59%

Table 1: Compression results for random text files.