UTAH STATE UNIVERSITY
ELECTRICAL ENGINEERING DEPARTMENT

EE 7680 Information Theory

Source Coding Programming Assignment

On the website for the course, `http://www.engineering.usu.edu/classes/ece/7680`, under "Programming assignments" is a large ($>$ 1 million character) file named `testdata.txt` consisting of the A – Z, 0 – 9, and space. Download the file into a directory where you can work on it.

Characterization of the source. Determine the source probabilities for the 37 symbols employed in the file.

Huffman Coding Devise a Huffman code suitable for this source. Determine the average codeword length for this code, based on the statistics of the data file. Determine how long the compressed file would be if coded using this Huffman code.

Lempel-Ziv coding Using `gzip`, `compress`, `pkzip`, or other popular Lempel-Ziv type stream compression algorithm, compress the data file.

Arithmetic Coding Now the heart of the assignment. Implement an arithmetic coder and compress the data file using your arithmetic coder.

Comparison Compare and discuss your compression results. (Provide numbers, tables, figures, etc., as appropriate.)