Project Proposal

Our group project is about compression, and we assume to use three different algorithms to implement this function, the run-length, the LZW and the Burrows-Wheeler. Zhe Wang is assigned to the first algorithm, Xining Li is assigned to the second one, and Laiquan Tang is assigned to the last one.

@Zhe Wang

Run-length encoding is a very simple form of lossless data compression in which runs of data (that is, sequences in which the same data value occurs in many consecutive data elements) are stored as a single data value and count, rather than as the original run. This is most useful on data that contains many such runs. For example, a set of data string “AAAABBBCCDEEEE”, after run-length encoding the data can be compressed into 4A3B2C1D4E(change 14 units into 10 units).The algorithm is using a temporary function Q read the first data, then compare the next data with Q, if the data is the same, the counter plus 1;If it is different, then output Q and the value in the counter and initial the counter to 0, the value Q change to the next data. Repeat this to complete the data compression

@Xining Li

The scenario described by Welch's encodes sequences of 8-bit data as fixed-length 12-bit codes. The codes from 0 to 255 represent 1-character sequences consisting of the corresponding 8-bit character, and the codes 256 through 4095 are created in a dictionary for sequences encountered in the data as it is encoded. At each stage in compression, input bytes are gathered into a sequence until the next character would make a sequence for which there is no code yet in the dictionary. The code for the sequence (without that character) is added to the output, and a new code (for the sequence with that character) is added to the dictionary.

The idea was adapted to other situations. For a reduced alphabet, the full 12-bit codes yielded poor compression unless the image was large, so the idea of a variable-width code was introduced: codes typically start one bit wider than the symbols being encoded, and as each code size is used up, the code width increases by 1 bit, up to some prescribed maximum (typically 12 bits). When the maximum code value is reached, encoding proceeds using the existing table, but new codes are not generated for addition to the table.

In this project, I will write the compress and decompress functions to achieve this algorithms.

@Laiquan Tang

The Burrows-Wheeler Transform (also known as Block-Sorting) is at the base of compression algorithms which are the state of the art in lossless data compression.

In this project, I will write codes to achieve this algorithms and analyze the algorithms which use this technique.

Further the more, we will compare the advantages and disadvantages of those three different algorithms.