Programming Practicum (CS571)

Date: 6 Sep 22

Assignment-3

Q.1 **Variable length coding:** The task involves encoding each Latin alphabet and non-printable characters present in a piece of text into a binary string and decoding the complete binary string to recover the original text. To accomplish this, write programs for the following subtasks.

1. Parse the text provided and find the number of occurrences of each Latin alphabet. Print it in a tabular form, where the first column lists alphabets and the second their respective number of occurrences.
2. Sort the table in descending order of the number of occurrences of alphabets.
3. Implement a fixed length encoding algorithm to represent each alphabet as a binary string.
4. Decode the bit stream corresponding to the fixed length coding and recover the original text.
5. Implement the variable length coding algorithm given here and present it in a tabular structure. Note that the length of binary codes corresponding to Latin alphabets depends on the frequency of their occurrence in the text.
6. Decode the bit stream corresponding to the variable-length coding and recover the original text.
7. Compare the number of bits in the encoded string generated using 3 and 5.
8. Compare the lengths of fixed-length and variable-length codes for each alphabet by printing them next to each other in a tabular format. Also, in the same tabular format, print their ratio (variable-length/fixed length), the product of the ratio with the number of occurrences in the text for each alphabet.

TextToBeEncoded = "*Variable-length codes can allow sources to be compressed and decompressed with zero error (lossless data compression) and still be read back symbol by symbol. With the right coding strategy an independent and identically-distributed source may be compressed almost arbitrarily close to its entropy. This is in contrast to fixed length coding methods, for which data compression is only possible for large blocks of data, and any compression beyond the logarithm of the total number of possibilities comes with a finite (though perhaps arbitrarily small) probability of failure*"

Source: https://en.wikipedia.org/wiki/Variable_length_coding