Abstract

In this paper, a new text transformation technique called Semi-Adaptive Substitution Coder for Lossless Text Compression is proposed. The rapid advantage of this Substitution Coder is that it substitutes the codewords by referring the reference of the word’s position in the dictionary to expedite the dictionary mapping and also codewords are shorter than words and, thus, the same amount of text will require less space. In general, text transformation needs an external dictionary to store the frequently used words. To preserve this transformation method in a healthy way, a semi-adaptive dictionary is used and therefore which reduces the expenditure of memory overhead and speeds up the transformation because of the smaller size dictionary. This new transformation algorithm is implemented and tested using Calgary Corpus and Large Corpus. In this implementation Semi-Adaptive Substitution Coder in connection with a popular bzip2 and commonly used Gzip compressors improve the compression performance by about 7–9% on large files.

References
Semi-Adaptive Substitution Coder for Lossless Text Compression

- Robert Franceschini, Amar Mukherjee, "Data Compression Using Encrypted
Semi-Adaptive Substitution Coder for Lossless Text Compression


Index Terms

Computer Science Information Sciences

Keywords

Transformation preprocessing adaptive dictionary compression decompression.