An Efficient Text Compression for Massive Volume of Data

Abstract

To propose a new text compression technique for ASCII texts for the purpose of obtaining good performance on various document sizes. This algorithm is composed of two stages. In the
first stage, the input strings are converted into the dictionary based compression. In the second stage, the redundancy of the dictionary based compression is reduced by Burrows wheeler transforms and Run length coding. The algorithm has good compression ratio and reduces bit rate to execute the text with increase in the speed.

Reference

- Jirapond Tadrat and Veera Boonjing, 2008"An Experiment study on Transformation for Compression using stop lists and Frequent words” IEEE Transactions on information technology.
- Data compression: the complete reference By David Salomon
- Compression test results, corpus.canterbury.ac.nz/

Index Terms

Computer Science  Information Theory

Key words

Dictionary Based Encoding (DBE)  Burrows-Wheeler Transform
(BWT)
Run Length
Encoding (RLE).