

Lossless Text Compression Technique Using Syllable Based Morphology

[Akman, I](#) (Akman, Ibrahim)^[1]; [Bayindir, H](#) (Bayindir, Hakan)^[1]; [Ozleme, S](#) (Ozleme, Serkan)^[1]; [Akin, Z](#) (Akin, Zehra); [Misra, S](#) (Misra, Sanjay)^[1]

Abstract

In this paper, we present a new lossless text compression technique which utilizes syllable-based morphology of multi-syllabic languages. The proposed algorithm is designed to partition words into its syllables and then to produce their shorter bit representations for compression. The method has six main components namely source file, filtering unit, syllable unit, compression unit, dictionary file and target file. The number of bits in coding syllables depends on the number of entries in the dictionary file. The proposed algorithm is implemented and tested using 20 different texts of different lengths collected from different fields. The results indicated a compression of up to 43%.