

# An adaptive character wordlength algorithm for data compression

Hussein Al-Bahadila<sub>a, \_</sub>, Shakir M. Hussain<sub>b</sub>

<sup>a</sup> Department of Computer Science, Amman Arab University for Graduate Studies, Amman, Jordan

<sup>b</sup> Department of Computer Science, Applied Science University, Amman, Jordan

Received 15 May 2006; accepted 16 May 2007

## Abstract

This paper presents a new and efficient data compression algorithm, namely, the adaptive character wordlength (ACW) algorithm, which can be used as complementary algorithm to statistical compression techniques. In such techniques, the characters in the source file are converted to a binary code, where the most common characters in the file have the shortest binary codes, and the least common have the longest; the binary codes are generated based on the estimated probability of the character within the file. Then, the binary coded file is compressed using 8 bits character wordlength. In this new algorithm, an optimum character wordlength,  $b$ , is calculated, where  $b > 8$ , so that the compression ratio is increased by a factor of  $b/8$ . In order to validate this algorithm, it is used as a complement algorithm to Huffman code to compress a source file having 10 characters with different probabilities, and these characters are randomly distributed within the source file. The results obtained and the factors that affect the optimum value of  $b$  are discussed, and, finally, conclusions are presented.

**Keywords:** Data compression; Coding; Huffman coding; Binary code