

Burrows-Wheeler Text Compression with Fountain Codes

Bertrand Ndzana Ndzana^{*}, Amin Shokrollahi^{*}, Jürgen Abel[†]
^{*}EPFL, [†]Ingenieurbüro Dr. Abel GmbH - Lechstrasse 1 - Germany

An algorithm in a purely lossless text compression setting based on Fountain Codes and the Burrows-Wheeler Transform (BWT) is presented. The proposed scheme consists of five stages. The first stage, the BWT performs a permutation of the input symbols, which is the basis for the following stages. The second stage, a Run Length Encoding replaces all runs, which have a length of two or more symbols, by a run consisting of exactly two symbols. The third stage transforms the local structure of the BWT output stream into an index stream with a global structure and is called a Global Structure Transform (GST). The last two stages are the Entropy Coder (EC) and Fountain Coder (FC). The length of a run is transmitted into a separate data stream with an EC. The GST-index stream is transmitted with a layered FC. The FC stage requires two passes. First the statistical modeling part to learn the probability distributions and the second pass to accomplish the encoding. The FC's modeling part follows the uniform segmentation [1]. The FC's encoding part follows the Closed-Loop Iterative Doping algorithm together with the multilevel stage decoding Belief Propagation. Our algorithm offers encouraging compression rate performance for large files. The results of the proposed scheme and other compression schemes are summarized in the following tables. *IFOUNT05* is the approach presented in this paper, where the GST stage is the Incremental Frequency Count [2].

File	Size	IFOUNT05	GZIP-B	BZIP-9	PPMD5
<i>Compression rates for the Calgary Corpus</i>					
Avg./bps	369536	3.21	3.24	2.64	2.65
<i>Compression rates for the Large Canterbury Corpus</i>					
Avg./bps	3719828	1.89	2.30	1.79	1.70

References

- [1] Haixiao Cai, Kulkarni S.R., Verdu S., “Universal entropy estimation via block sorting”, *IEEE Transactions on Information Theory*, Volume 50, Issue 7, July 2004 Page(s):1551 - 1561
- [2] Abel J., “A fast and efficient post BWT-stage for the Burrows-Wheeler Compression Algorithm”, *Proceedings of the IEEE Data Compression Conference 2005, Snowbird, Utah*, J. A. Storer and M. Cohn, Eds., 449, 2005.