

Author(s): Ferreira, AJ (Ferreira, Artur J.); Oliveira, AL (Oliveira, Arlindo L.); Figueiredo, MAT (Figueiredo, Mario A. T.)

Editor(s): Filipe, J; Obaidat, MS

Title: On the Suitability of Suffix Arrays for Lempel-Ziv Data Compression

Source: E-Business and Telecommunications, 48: 267-280 2009

Book series title: Communications in Computer and Information Science

Language: English

Document Type: Proceedings Paper

Conference Title: 5th International Conference on e-Business and Telecommunication

Conference Date: JUL 26-29, 2008

Conference Location: Oporto, PORTUGAL

KeyWords Plus: Construction; Algorithm

Abstract: Lossless compression algorithms of the Lempel-Ziv (LZ) family are widely used nowadays. Regarding time and memory requirements, LZ encoding is much more demanding than decoding. In order to speed up the encoding process, efficient data structures, like suffix trees, have been used. In this paper, we explore the use of suffix arrays to hold the dictionary of the LZ encoder, and propose an algorithm to search over it. We show that the resulting encoder attains roughly the same compression ratios as those based on suffix trees. However, the amount of memory required by the suffix array is fixed, and much lower than the variable amount of memory used by encoders based on suffix trees (which depends on the text to encode). We conclude that suffix arrays, when compared to suffix trees in terms of the trade-off among time, memory, and compression ratio, may be preferable in scenarios (e.g., embedded systems) where memory is at a premium and high speed is not critical.

Addresses: [Ferreira, Artur J.] Inst Super Engn Lisboa, Lisbon, Portugal

Reprint Address: Ferreira, AJ, Inst Super Engn Lisboa, Lisbon, Portugal.

E-mail Address: arturj@cc.isel.ipl.pt; aml@inesc-id.pt; mario.figueiredo@lx.it.pt

Publisher: SPRINGER-VERLAG BERLIN

Publisher Address: HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY

ISSN: 1865-0929

ISBN: 978-3-642-05196-8

ISI Document Delivery No.: BPV35