

Accepted Manuscript

A survey of string orderings and their application to the Burrows-Wheeler transform

Jacqueline W. Daykin, Richard Groult, Yannick Guesnet, Thierry Lecroq, Arnaud Lefebvre et al.

PII: S0304-3975(17)30154-8
DOI: <http://dx.doi.org/10.1016/j.tcs.2017.02.021>
Reference: TCS 11090

To appear in: *Theoretical Computer Science*

Received date: 26 June 2016
Revised date: 29 September 2016
Accepted date: 14 February 2017

Please cite this article in press as: J.W. Daykin et al., A survey of string orderings and their application to the Burrows-Wheeler transform, *Theoret. Comput. Sci.* (2017), <http://dx.doi.org/10.1016/j.tcs.2017.02.021>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



A Survey of String Orderings and Their Application to the Burrows-Wheeler Transform [☆]

Jacqueline W. Daykin^{a,b,d}, Richard Groult^{c,d}, Yannick Guesnet^d, Thierry Lecroq^d, Arnaud Lefebvre^d, Martine Léonard^d, Élise Prieur-Gaston^d

^a*Department of Computer Science, Aberystwyth University (Mauritius Branch Campus), Quartier Militaire, Mauritius*

^b*Department of Computer Science, Royal Holloway, University of London, UK*

^c*Modélisation, Information et Systèmes (MIS), Université de Picardie Jules Verne, Amiens, France*

^d*Normandie Univ., UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, 76000 Rouen, France*

Abstract

For over 20 years the data clustering properties and applications of the efficient Burrows-Wheeler transform have been researched. Lexicographic suffix-sorting is induced during the transformation, and more recently a new direction has considered alternative ordering strategies for suffix arrays and thus the transforms. In this survey we look at these distinctly ordered bijective and linear transforms. For arbitrary alphabets we discuss the V -BWT derived from V -order and the D -BWT based on lex-extension order. The binary case yields a pair of transforms, the binary Rouen B -BWT, defined using binary block order. Lyndon words are relevant to implementing the original transform; the new transforms are defined for analogous structures: V -words, indeterminate Lyndon words, and B -words, respectively. There is plenty of scope for further non-lexicographic transforms as indicated in the conclusion.

Keywords: algorithm, bijective, binary alphabet, block order, Burrows-Wheeler Transform, B -word, data clustering, degenerate, GB -word, generic alphabet, generic block order, indeterminate Lyndon word, inverse transform, lexicographic order, linear, Lyndon word, string, suffix array, suffix-sorting, T -order, V -order, word.

1. Introduction

In this survey we review recent advances in schemes for computing permutations of texts in the context of the Burrows-Wheeler transform (BWT) [7]. Considerable interest in this transform, spanning over 20 years, ensues from its

[☆] *This paper is in honour of the esteemed stringologist Professor Costas S. Iliopoulos in recognition of his significant contributions to theoretical computer science and stringology.*

Download English Version:

<https://daneshyari.com/en/article/6875668>

Download Persian Version:

<https://daneshyari.com/article/6875668>

[Daneshyari.com](https://daneshyari.com)