Journal of Automata, Languages and Combinatorics **21** (2016) 3, 131 – 147

ENHANCED COVERS OF REGULAR AND INDETERMINATE STRINGS USING PREFIX TABLES

Ali Alatabbi[(A)] Abu Sayed Md. Sohidull Islam[(B)]
Mohammad Sohel Rahman[(C,F)] Jamie Simpson[(D)] W. F. Smyth[(B,E,G)]

[(A)] *Department of Informatics, King's College London, Strand, London, WC2R 2LS, UK*
ali.alatabbi@kcl.ac.uk

[(B)] *Algorithms Research Group, Department of Computing & Software, McMaster University, Hamilton ON L8S 4K1, Canada*
sohansayed@gmail.com, smyth@mcmaster.ca

[(C)] *AℓEDA Group, Department of Computer Science and Engineering (CSE), Bangladesh University of Engineering and Technology (BUET)*
msrahman@cse.buet.ac.bd

[(D)] *Department of Mathematics and Statistics, Curtin University of Technology, Bentley, Western Australia 6102, Australia*
Jamie.Simpson@curtin.edu.au

[(E)] *Centre for Combinatorics of Words and Applications, Mathematics Department, Murdoch University, 90 South Street, Murdoch, Western Australia 6150, Australia*

ABSTRACT

A ***cover*** of a string $\boldsymbol{x} = \boldsymbol{x}[1..n]$ is a proper substring $\boldsymbol{u}$ of $\boldsymbol{x}$ such that $\boldsymbol{x}$ can be constructed from possibly overlapping instances of $\boldsymbol{u}$. A recent paper [12] relaxes this definition — an ***enhanced cover*** $\boldsymbol{u}$ of $\boldsymbol{x}$ is a border of $\boldsymbol{x}$ (that is, a proper prefix that is also a suffix) that covers a *maximum* number of positions in $\boldsymbol{x}$ (not necessarily all) — and proposes efficient algorithms for the computation of enhanced covers. These algorithms depend on the prior computation of the ***border array*** $\boldsymbol{\beta}[1..n]$, where $\boldsymbol{\beta}[i]$ is the length of the longest border of $\boldsymbol{x}[1..i]$, $1 \leq i \leq n$. In this paper, we first show how to compute enhanced covers using instead the ***prefix table***: an array $\boldsymbol{\pi}[1..n]$ such that $\boldsymbol{\pi}[i]$ is the length of the longest substring of $\boldsymbol{x}$ beginning at position i that matches a prefix of $\boldsymbol{x}$. Unlike the border array, the prefix table is robust: its properties hold also for ***indeterminate strings*** — that is, strings defined on *subsets* of the alphabet Σ rather than individual elements of Σ. Thus, our algorithms, in addition to being more space-efficient than those of [12], allow us to easily extend the computation of enhanced covers to indeterminate strings. Both for regular and indeterminate strings, our algorithms execute in expected linear time and our experimental results suggest that they are faster than those of [12] in practice. Along the way we establish an important theoretical result: that the expected maximum length of any border of any prefix of a regular string $\boldsymbol{x}$ is approximately 1.64 for binary alphabets, less for larger ones.

Keywords: string, word, indeterminate string, prefix table, border array, cover, enhanced cover, algorithm

---

[(F)] Currently on a sabbatical leave from BUET. Part of this research was carried out when Rahman was visiting McMaster University under an ACU Titular Fellowship.

[(G)] Supported in part by a grant from the Natural Sciences & Engineering Research Council (NSERC) of Canada.