Louza, Felipe A.; Mhaskar, Neerja; Smyth, W. F.
A new approach to regular & indeterminate strings. (English) Zbl 1477.68555

Summary: In this paper we propose a new, more appropriate definition of regular and indeterminate strings. A regular string is one that is “isomorphic” to a string whose entries all consist of a single letter, but which nevertheless may itself include entries containing multiple letters. A string that is not regular is said to be indeterminate. We begin by proposing a new model for the representation of strings, regular or indeterminate, then go on to describe a linear time algorithm to determine whether or not a string $x = x[1..n]$ is regular and, if so, to replace it by a lexicographically least (lex-least) string $y$ whose entries are all single letters. Furthermore, we connect the regularity of a string to the transitive closure problem on a graph, which in our special case can be efficiently solved. We then introduce the idea of a feasible palindrome array MP of a string, and prove that every feasible MP corresponds to some (regular or indeterminate) string. We describe an algorithm that constructs a string $x$ corresponding to given feasible MP, while ensuring that whenever possible $x$ is regular and if so, then lex-least. A final section outlines new research directions suggested by this changed perspective on regular and indeterminate strings.

MSC:
68W32 Algorithms on strings
68R15 Combinatorics on words

Keywords:
indeterminate string; degenerate string; generalized string; transitive closure; palindrome; maximal palindrome array; Manacher’s condition; reverse engineering; algorithm; stringology

Software:
Algorithm 97

Full Text: DOI arXiv

References:
[26] F. Gray, Pulse code communication.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.