Journal of Automata, Languages and Combinatorics **27** (2022) 1–3, 187 – 197

NONDETERMINISTIC STATE COMPLEXITY OF SITE-DIRECTED INSERTION

Oliver A. S. Lyon[(A)] Kai Salomaa[(A,B)]

[(A)] *School of Computing, Queen's University*
Kingston, Ontario K7L 2N8, Canada
oliver.lyon@queensu.ca salomaa@queensu.ca

ABSTRACT

Site-directed insertion (a. k. a. outfix-guided insertion) is a controlled insertion operation where an outfix of the inserted string has to match a substring of the target string. We show that if L_1 and L_2 have NFAs (nondeterministic finite automata) with N and M states, respectively, the site-directed insertion of L_2 into L_1 can be recognized by an NFA with $3NM$ states. This improves the known upper bound for nondeterministic state complexity by an additive factor of $2N$. As our main result we establish for the nondeterministic state complexity of site-directed insertion a lower bound $3NM - M$.

Keywords: finite automaton, nondeterministic state complexity, fooling set, bio-inspired language operations

1. Introduction

Gene insertion and deletion are basic operations occurring in DNA recombination. *Site-directed mutagenesis* is one of the most important laboratory techniques for generating mutations on specific sites of DNA using polymerase chain reaction [16, 17]. Since the insertions occurring in DNA strands depend on the context, Kari and Thierrin [15] modeled such bio-operations as contextual insertions [18]. Later contextual insertion-deletion systems have been studied, e. g., by Daley et al. [7] and Takahara and Yokomori [20]. Site-directed insertion, a. k. a. outfix-guided insertion [4, 5] is another biologically motivated context-dependent insertion operation.

Site-directed insertion is an overlapping insertion operation, where a nontrivial outfix of the inserted string has to match a substring of the target string. More formally, the site-directed insertion of a string y into a string x consists of all strings x_1uzvx_2 where $x = x_1uvx_2$, $y = uzv$ and u, v are nonempty strings. The definition requires that an outfix of the inserted string must match a substring of the target string. Site-directed insertion relates to the non-overlapping insertion operation analogously as the overlap assembly or chop operations [6, 8, 11, 12] relate to concatenation.

[(B)] Supported by the Natural Sciences and Engineering Research Council of Canada Grant OGP0147224.