Journal of Automata, Languages and Combinatorics **2** (1997) 4, 227–268

RANKING AND UNRANKING OF LEXICOGRAPHICALLY ORDERED WORDS: AN AVERAGE-CASE ANALYSIS

JENS LIEBEHENSCHEL
Fachbereich Informatik, Johann Wolfgang Goethe-Universität
D-60054 Frankfurt am Main, Germany
e-mail: jens@sads.informatik.uni-frankfurt.de

ABSTRACT

We consider all words of length n of a formal language. If these words are arranged according to the lexicographical order, then ranking means to determine the position of a word of the language. Unranking is the inverse operation of ranking. For a given formal language we compute the average length of the shortest prefix of a word to be read to determine its position, if the word is read from left to right. The length of the shortest prefix to be read depends on the language only, not on the ranking algorithm. After having derived a general expression, we demonstrate the result by discussing various concrete applications. Ranking will be analyzed for regular languages, permutations, subsets of sets, the DYCK language, the MOTZKIN language, extended ordered binary trees according to ZAKS, ordered binary trees according to ER, extended ordered t-ary trees according to RUSKEY, ordered trees with bounded height and CAYLEY trees. Furthermore, ranking and unranking algorithms will be presented for the languages aforementioned.

Keywords: ranking, unranking, lexicographical order, average-case analysis, regular languages, permutations, subsets, Dyck language, Motzkin language, trees.

1. Introduction

A number of papers dealing with the lexicographical generation as well as with the ranking and unranking functions of combinatorial objects such as permutations and classes of trees were published. In [10] an average-case analysis of a frequently used method to generate words lexicographically has been presented. This method generates the successor of a word by reading the word from right to left until the suffix to be changed is found. Note that the average length of the suffix to be read in order to generate the successor of a word depends on the language only, not on the algorithm.

In the second section of this paper we develop an expression for the average length of the prefix of a word to be read by the ranking algorithm, i. e. the algorithm which computes the position of the word according to the lexicographical order. This expression depends on the language only, not on the algorithm. The strategy employed below is to compute the number of lexicographically smaller words by reading the