Pattern Matching on Weighted Strings
(*Abstract*)

Jakub Radoszewski*

Institute of Informatics, University of Warsaw, Warsaw, Poland
Banacha 2, 02-097 Warszawa, Poland
jrad@mimuw.edu.pl

Keywords: weighted string, weighted sequence, uncertain string, Position-Weight Matrix (PWM), pattern matching

A weighted string is a sequence of probability distributions over a given finite alphabet Σ. A weighted string represents many standard strings, each with the probability of occurrence equal to the product of probabilities of its letters at subsequent positions of the weighted string. Usually a threshold $1/z > 0$ is specified and one considers as matches only the strings for which the probability of occurrence is at least $1/z$.

Weighted strings, also known as position weight matrices or uncertain strings, arise naturally in many applications. In molecular biology, position weight matrices were introduced as an alternative to consensus sequences and may appear due to flexible sequence modeling, such as binding profiles of molecular sequences. Weighted strings are also present in mining applications due to imprecise data measurements or when observations are private and thus sequences of observations may have artificial uncertainty introduced deliberately. Weighted strings can also be viewed as a generalization of indeterminate strings (i.e., degenerate strings).

This talk will focus on the solutions to Weighted Pattern Matching problem, in which we are to find all positions of the given weighted text where a given string pattern occurs with probability above the threshold, and its indexing variant. We will also survey other algorithmic results on weighted strings, including a variant of Weighted Pattern Matching in which both the text and the pattern are weighted, practical approaches to Weighted Pattern Matching, and differences between the longest common subsequence and shortest common supersequence problems on weighted strings.

* The author was supported by the "Algorithms for text processing with errors and uncertainties" project carried out within the HOMING programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund.

Jakub Radoszewski: Pattern Matching on Weighted Strings, p. 1.
Proceedings of PSC 2019, Jan Holub and Jan Žďárek (Eds.), ISBN 978-80-01-06618-8 © Czech Technical University in Prague, Czech Republic