SEARCHING FOR UNIQUE DNA SEQUENCES WITH THE BURROWS-WHEELER TRANSFORM

Rafał Pokrzywa

Department of Computer Science, Silesian University of Technology, Gliwice, Poland

The objective of this study was to present an efficient algorithm that effectively aids the problem of searching for unique DNA sequences in the set of genes. The presented algorithm is based on the Burrows-Wheeler Transform (BWT), a very fast and effective data compression algorithm. The developed algorithm exploits all the advantages offered by the BWT algorithm and the suffix array data structure. It allows obtaining a structure that is ideal for solving many problems related to the pattern-matching problem. This algorithm is applicable to the identification of yeast species as well as to many other computational molecular biology problems like searching for repetitive structures in genomic sequences, designing of DNA hybridization probes and many more.

Keywords: Burrows-Wheeler Transform, yeasts species, suffix array, longest common prefix