

Parameterized String Matching Algorithms with Application to Molecular Biology

R Prasad, S Agarwal, S Misra

Abstract

In the molecular biology, it is said that two biological sequences tend to have similar properties if they have similar 3-D structures. Hence, it is very important to find not only similar sequences in the string sense, but also structurally similar sequences from the database. Parameterized string matching has been used to find structurally similar sequences from the database. In the parameterized string matching problem, a given pattern P is said to match with a sub-string t of the text T , if there exist a bijection from the symbols of P to the symbols of t . Salmela and Tarhio solve the parameterized string matching problem in sub-linear time by applying the concept of q -gram in the Horspool algorithm (FPBMH). In this paper, we extend the Boyer Moore type algorithms: Smith, Raita and Quick Search, to solve the same problem by using the q -gram. We compare the performance of: FPBMH, Smith, Raita, and Quick search algorithms on DNA alphabet and found that Smith algorithm perform better than FPBMH algorithm.

Keywords: Algorithm, prev-encoding, parameterized matching, molecular biology, and RGF String.

Full Text:

[EMAIL FULL TEXT](#) 
[DOWNLOAD FULL TEXT](#) 