

Parameterized String Matching Algorithms with Application to Molecular Biology.

Rajesh Prasad¹, Suneeta Agarwal² and Sanjay Misra³.

^{1,2}Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology, Allahabad, 211004, INDIA.

³Department of Department of Computer Engineering, School of Information and Communication Technology, Federal University of Technology, Minna, Nigeria.

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.

Email: rajesh_ucer@yahoo.com, suneeta@mnnit.ac.in, smisra@futminna.edu.ng

Received 18th October, 2010

Accepted 16th December, 2010