Abstract

String matching is to find all the occurrences of a given pattern in a large text both being sequence of characters drawn from finite alphabet set. This problem is fundamental in computer Science and is the basic need of many applications such as text retrieval, symbol manipulation, computational biology, data mining, and network security. Bit parallelism method is used for increasing the processing speed of String matching algorithm. Standard Shift OR algorithm is used to perform approximate string matching. The algorithm is a filter which finds out false matches besides detecting correct matches. To improve the efficiency of basic Shift OR algorithm by reducing the number of false matches that is detected along with the correct matches by the algorithm, proposed Shift OR with consecutive q grams has been implemented. In the algorithm instead of reading a single character at a time, it read q characters at once. Extensive experiments have been done with the algorithm and the results are compared with basic version of shift OR algorithms. The number of false matches also reduced considerably. The gain is due to the improved filtering efficiency caused by q-grams.
Improved Approximate Multiple Pattern String Matching using Consecutive Q Grams of Pattern


Index Terms

Computer Science  Pattern Recognition
Keywords
Approximate String matching  Bit parallelism  Shift OR String Matching  q Grams