

GENOMICS PAPER

Simple and Efficient Pattern Matching Algorithms for Biological Sequences

Zeyad elsayed elkomy rashed



Abstract:

The growth of biological data is pushing us to discover solutions in many areas of computational bioinformatics. In different stages of the computational pipelines, pattern matching is a very practical operation. For example, pattern matching enables us to find the locations of particular DNA sequences or subsequence in a database. Furthermore, in these expanding biological databases, some patterns are updated over time. To perform faster searches, high-speed pattern matching algorithms are needed. In paper we introduce pattern matching algorithms that are used to speed up search on large DNA sequences. The proposed algorithms raise performance by utilizing word processing (instead of the character processing) and also by searching the least frequent word of the pattern in the sequence. The experimental results showed that these algorithms are distinguished over other algorithms in terms of Speed and accuracy.

INTRODUCTION:

In the pattern matching problem, a sequence is scanned to detect the locations of it. It is important to address these problems due to their applications in various fields such as image and text processing, search engines, and information retrieval. Notably, the pattern matching problem arises in different scopes of computational bioinformatics, example biomarker discovery, sequence alignment. In these disciplines, there is a need to recognize the locations of multiple patterns, including those of amino acids and nucleotides in Databases and we use it in different fields. Although these algorithms are efficient the development of it is still required. because many current algorithms may not be well scalable for databases or large DNA sequences due to high-computational costs. the current paper focuses on the exact pattern matching problem which finds all the occurrences of a pattern in a text by using algorithms to Repair previous defects and it is divided into a Different stages. Nowadays, the computational length of almost all processors is 32 or 64 bits in each execution cycle so they can process 4 or 8 bytes of data instantly. Therefore, 4 or 8 characters, indicated by a word. we introduces also an algorithms to conduct word-based comparisons. The word processing is performed by utilizing power of the processor. This approach creates a new class of string-matching algorithms that improve the performance of character-based algorithms. As a result, the performance improves in terms of time cost.