Research , paper about BFM: a forward backward string matching algorithm with improved shifting for information retrieval

Assignment) Abtract:

Extraction data and gain valuable information from given text is an important part of data mining . Nowadays,applying a string matching algorithm is very necessary As a result of the growing files' sharing over the Internet and the urgent need for a cloud computing .in this paper, we are gonna discuss a new algorithm named Back and Forth Matching algorithm (BFM),which decreases the number of characters comparisons .It makes an improvement than other algorithms that works faster in alignment process , by matching pattern from both forward and backward direction with text. So it simplifies searching of large string to match with pattern.

Introduction :

To find a required text (pattern)in a text decuoment is a simple task.but it may be take long time and huge file size. As a result, a fast pattern matching algorithm is an essential component of the page ranking in search engine, checking syntax ,spelling mistakes and many applications .it also important in bioinformatics, DNA sequences matching, and behavior analysis as well. To search through the text, the pattern is used as a window of length m that moves over the text T starting from it's leftmost character. The target of the algorithm is to match the characters of the window with the text characters in multiple attempts . This succession of attempts and window shifts continues, until the right end of the window reaches the length of the text. This type of window mechanism is known as the sliding window mechanism. The aim of these algorithms are to minimize the character comparisons and to maximize the length of shifts . we've presented a new string-matching algorithm BFM. More importantly, the algorithm does both forward and backward checking and decreases the number of efforts required to match the window with the text during the matching phase.

ASSIGNMENT 2: Related Work ;

The Boyer Moore algorithm is one of the most renowned, efficient and extensively used pattern matching algorithm. It preprocesses the pattern and, in the event of a mismatch, calculates the pattern's maximal changes using two heuristics(the good and bad heuristics). Boyer Moore chooses the better of the two heuristics, thus can slide the pattern by maximum number of characters.

It descripes another way to improve the algorithm by shifting. Another algorithm is called the Quick Search algorithm. it proposed a combination of Quick Search and the Skip Search algorithm. hybrid algorithm was proposed using the idea of Quick-Skip and Boyer–Moore algorithms.

The Knuth–Morris–Pratt (KMP) algorithm, which operates similarly to a naive algorithm, is another highly used algorithm.

instead of checking all the characters after each shift, the algorithm preprocesses P to build a table. Proposes a modern algorithm that combines the Boyer–Moore and KMP ideas.

Beside all these algorithms, a comperatively old string matching technique is Rabin Carp string searching algorithm that can search both single and multiple patterns in a string, that deploys a modified version of the Rabin Carp algorithm using a GPU processor.