New and Efficient Recursive-based String Matching Algorithm (RSMA-FLFC)

Abstract

The need for simple and efficient string matching algorithms is essential for many applications, and especially for database query. In this paper, two major algorithms are proposed, namely first least frequency character algorithm (FLFC) and recursive-based string matching algorithm (RSMA). FLFC is considered as an enhanced version of scan for lowest frequency character SLFC proposed by Horspool [12]. FLFC algorithm extracts first least frequency character in the pattern and identifies the occurrences of such character in the whole text in a preprocessing phase, while the recursive algorithm (RSMA) recursively partitioning the pattern and the targeted substring in the text and compares them at mid-point (q) each time. The FLFC search accelerates the searching process, while RSMA enhances the speed of performance of the matching phase. The extensive testing and comparisons with Naïve (Brute force), Boyer-Moore (BM), and the FLFC without deploying recursive matching show that the proposed algorithms enhance the speed of performance dramatically.

References

- Ukkonen, E. 1985. Algorithms for approximate string matching. Information and Control, 64(1-3) 100–118

**Index Terms**

Computer Science

Algorithms

**Keywords**
Recursive string matching  brute force  Boyer-Moore  least frequency characters