

FAST-SEARCH ALGORITHMS: NEW EFFICIENT VARIANTS OF THE BOYER-MOORE PATTERN-MATCHING ALGORITHM¹

DOMENICO CANTONE and SIMONE FARO

*Department of Mathematics and Computer Science, University of Catania
Viale A. Doria n. 6 95125, Catania, Italy
e-mail: {cantone, faro}@dmi.unict.it*

ABSTRACT

We present two variants of the Boyer-Moore string matching algorithm, named *Fast-Search* and *Forward-Fast-Search*, and compare them with some of the most effective string matching algorithms, such as Horspool, Quick Search, Tuned Boyer-Moore, Reverse Factor, and Berry-Ravindran. All algorithms are compared in terms of their run-time efficiency, number of text character inspections, and number of character comparisons.

It turns out that the new proposed variants, though not linear, achieve very good results especially in the case of very short patterns or small alphabets.

Keywords: string matching, experimental algorithms, text processing, comparisons, efficient algorithms, text searching

1. Introduction

Given a text T and a pattern P over some alphabet Σ , the *string matching problem* consists in finding *all* occurrences of the pattern P in the text T . It is a very extensively studied problem in computer science, mainly due to its direct applications to such diverse areas as text, image and signal processing, speech analysis and recognition, information retrieval, computational biology and chemistry, etc.

Several string matching algorithms have been proposed over the years. The Boyer-Moore algorithm [1] deserves a special mention, since it has been particularly successful and has inspired much work. It is based upon three simple ideas: right-to-left scanning, bad character and good suffix heuristics. We will review it at length in Section 2.1. Many subsequent algorithms have been based on variations on how to apply the two mentioned heuristics.

Before entering into details, we need a bit of notations and terminology. A string P is represented as a finite array $P[0..m-1]$, with $m \geq 0$. In such a case we say

¹Full version of a submission presented at the *Prague Stringology Conference* (Czech Technical University in Prague, Czech Republic, September 22–24, 2003).