CHAPTER 1

INTRODUCTION

1.0 INTRODUCTION

Search algorithm is known as a universal problem-solving mechanism. It is an algorithm for searching an item (pattern) with a collection of item (text). Normally, it is a list of data documents that match the predefined matching criteria which is the outcome of search forms. This algorithm is use to seeking for more valuable information from available dataset in medical database. For example, search algorithm is needed when to search for word or phrase in a document, search for a student's name in a student list and search for approximate file name in a directory. There are several types of text searching algorithm which is Boyer-Moore algorithm (BM), Boyer Moore Horspool algorithm (BMH), Brute-Force algorithm (BF), Knuth-Morris Pratt algorithm (KMP), and Rabin-Karp algorithm (RK).
This research is concern on pattern matching problems. It is used to check the similarities between pattern matching and the text from database. In many field, such as computer engineering, computer science, bio-science database query and so on, text matching processing is essential and therefore applied frequently. To check whether this pattern is a substring of the text, the algorithm will compares a short string called pattern with a long string called text.

1.1 PROBLEM BACKGROUND

The most important problem in text searching is the exact text matching [1]. This can be describes in finding an exact related text to search an exact pattern input by user which expect the relevant output from their search in the shortest time. For example, there are given a text, $T$ of length $m$ and pattern, $P$ of length $n$ in matching text problem. The output from matching algorithm is either an indication that the pattern $P$ does not exist in $T$ or the starting index in $T$ is a substring matching $P$. Hence, want to find whether $P$ is a substring of $T$ starting from some index that matches pattern $P$. 
1.2 PROBLEM STATEMENT

Conventional method of search algorithm has raising some problems such as inefficiency of pattern search techniques. The problem consists of:

1.2.1 The amount of memory space needed by the computations.

1.2.2 The matching accuracy between pattern input and text in database.

1.3 OBJECTIVES

Based on the problems statement, the objectives of this study are:

i. To develop a searching function using Boyer-Moore Horspool algorithm based on keyword in a database.

ii. To provide an automatic search routine in any searching application.

iii. To evaluate performance of the proposed searching algorithm through character match percentage.