RETRIEVING TRANSLATED AL-QURAN DOCUMENTS USING EDIT DISTANCE TECHNIQUE

By

Ahmad Syukri Mohamad Zaid

Thesis Submitted in Fulfillment for the Degree of Bachelor Of Science

Faculty of Information Technology and Quantitative Science

MARA University of Technology

April 2001

ACKNOWLEDGMENTS

Putting out a research study this fast required a lot of help from people who went beyond the normal effort that goes into producing a research study to do everything that much more quickly.

I would like to take this opportunity to express my greatly appreciation and gratitude to my research supervisor, Prof. Madya Dr. Zainab Abu Bakar who has her support, supervision advice, guidance and also her suggestions on this dissertation. Without her guidance guideline, this research study may not be completed.

I would like to thank my course coordinator, Pn. Nor Elaiza Abd. Khalid and research examiner, Pn. Rosmah Abdul Latif. I gratefully acknowledge their precious time-spent undergoing in completing this research.

To my friends and colleagues in the Faculty of Information Technology and Quantitative Sciences and elsewhere, who are simply too numerous to mention, your time and support made my life much brighter.

Last, but not least, thanks to my parents, Mohamad Zaid and for helping me achieving a good education. They are the best parents in the world, continue to encourage and support me in every way. I am who I am because of both of them and I thank them for all that they have done for me.

ABSTRACT

This paper describes features and methods for string searching and comparison using Wagner and Fischer algorithm. The method is used in developing search engine because of its speed and efficiency since it is based on dynamic programming. The approach has two main sections: the methodology used in implementing, and the result. The methodology used in implementing contains the checking step, and the calculation of Dice coefficient. The result is the produced result from experiment. The method has been implemented as a C programming language. The Wagner and Fischer is used to retrieve documents from the translated Al-Quran. Quantitative and qualitative comparisons are made between the different Dice coefficients. The comparisons show that smaller Dice coefficient gives quantitative results but larger Dice coefficient produces qualitative results.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 Background

The string searching problem is to find all occurrences of a query word of in a translated Al-Quran documents. The problem is a fundamental concern of computer science. String searching methods are included in every text editor, word processor, search engine and many other applications. Generally, a string searching algorithm consists of two-step processing.

The first step is called checking step, and it is for determining whether the pattern matches with a substring of the documents. The other is called skipping step, and it is for detecting the next position in the text where the substring can possibly match with the pattern. The aim of a good string searching method is to decrease the number of character comparisons in the checking step and increase the jump in the skipping step. In order to find a complete match in the checking step, all characters in the query word must be compared with the corresponding characters in the documents. However, if one pair of characters is found to be unequal, the mismatch result is known. Obviously, different comparison orders yield different number of character comparisons. In this paper, a string searching method is discussed to produce a fast solution. The Wagner and Fischer algorithm make use of dynamic programming to get a good comparison order.