EVALUATION OF THE EFFECTIVENESS AND EFFICIENCY OF WAGNER AND FISCHER ALORITHM IN RETRIEVING HADITH TRANSLATED DOCUMENTS

KAMARUL ARIFIN B MUHAMAD

THESIS SUBMITTED IN PARTIAL FULFILMENT FOR THE DEGREE OF BACHELOR OF SCIENCE

FACULTY OF INFORMATION TECHNOLOGY
AND QUANTITATIVE SCIENCES
UNIVERSITI TEKNOLOGI MARA
SHAH ALAM

2002

ACKNOWLEDGEMENT

بسم الله الرحمن الرحيم
الحمد لله رب العالمين
الصلاة والسلام على أشرف الانبياء والمرسلين

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 Bte Abu Bakar who has her support, supervision advice, guidance and also her suggestion on this dissertation. Without her guidance guideline, this research study may not be completed. I would like to thank my thesis coordinator, Puan Nor Elaiza Abd Khalid and research examiner, Assoc Prof Nurazzah Bte Abd Rahman. 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.

My special appreciation is due to the Dean of the Faculty of Information Technology and Quantitative Sciences, Assoc Prof Azizi Ngah Tasir and the Programme Head of Bsc (Hons) Information Technology, Puan Jamaliah who is giving good management services throughout my degree program. My credits are

TABLE OF CONTENTS

ABSTRACT

Information technology has enabled information that can be in many forms such as text, image or sound, to be accessed widely using search terms via a computer. Due to this type of popularity and advanced in technology, there is an increased interest in searching Malay text to enable scholars and researchers to access the data on-line. This thesis studies the method of a string-searching algorithm. The studies describe features and methods for string searching and comparison using Wagner and Fischer algorithm. This thesis concerns a Malay language documents retrieval system. The method is used in developing search engine because of its speed and efficiency since it is based on dynamic programming. Dynamic programming is used to solve Edit Distance problem. 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. In this study, Wagner and Fischer algorithm, Hadith test collection that consists of Malay Hadith translation documents, query words and relevant judgements are used. The Wagner and Fischer are used to retrieve documents from translated Hadith. Before the experiments can be performed, the Hadith test collections must be built first. Two types of experiments are conducted. First experiment uses exact match, which is no method, is apply. Second experiment uses dynamic programmings apply with Wagner and Fischer algorithm.

CHAPTER 1

INTRODUCTION

1.1   Background

The study on information retrieval is how to determine and retrieve from a mass of prepared information, the part that is relevant to particular information needs (Sembok 1989). The main function of information retrieval system is to provide the users to perform searching effectively and efficiently.

In principle, information storage and retrieval is simple. Suppose there is a store of documents and a person formulates a question to the storekeeper to which the answer is a set of documents satisfying the information need expressed by his question. The storekeeper can obtain the set by reading all the documents in the store, retaining the relevant documents and discarding all the others. Although this solution seems to be perfect, but this it is obviously impracticable. Information retrieval is a study on how to determine and retrieve from a corpus of stored information; the part that is relevant to particular information needs (van Rijsbergen, 1979). The main function of information retrieval system is to provide the user with tools to perform searching effectively and efficiently.