INCORPORATING THESAURUS IN SEARCHING RELEVANT AYAT FROM THE TRANSLATED Al-QURAN

WAN IZNIF BIN WAN IBRAHIM

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

ACKNOWLEDGEMENTS

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

Praise to Allah the Almighty for providing me a superior knowledge throughout this period of study and for all that has bestowed on me. It is with His power the study is completed.

I would like to express a very special gratitude and thankfulness to my supervisor Assoc. Prof. Dr. Zainab Abu Bakar who is giving me reliance and shrewd guidance that have profited me very much. She who is also gave lots of support in this knowledge quest.

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 Bachelor (Hons.) Information Technology, Puan Jamaliah Taslim who are giving good management services throughout my degree program. My credits are also due to the computer laboratory staff of the faculty for their most generous assistance.

To my friends, Anuar Othman, Khairul Helmy Naburi, Razidi Abu, Mohd Zaini Abas, Mohd Rosmadi Mokhtar, and others in the faculty, who are simply too numerous to indicate, thanks for your time, understanding and buttress. May Allah will reciprocate you for all those things.

Last but not least, to my parents, and my siblings, Wan Izhan Wan Ibrahim and Wan Izlina Wan Ibrahim, I would like to forward my obliged to them for their continuous support during my study period, their patience and benevolence. Thank you.

والله علم بالصواب
أمين

CONTENT

ABSTRACT

This project incorporates the Rules Application Order (RAO) stemming algorithm with thesaurus approach by Rapizal. It is an opportunity to find out whether combining stemming with thesaurus will improve retrieval effectiveness and efficiency. Advances in information technology has made it possible for a wide range of text-based information to be searched and retrieved online, locally or from remote hosts. A wide range of text-based information is searched and retrieved from online connection to remote hosts or locally. This type of popularity is due to advancement in technology that is rapidly growing from day to day. There are many Malay word variants that have the same meaning available from Malay words. In order to overcome these word variant problems, the development of a computational technique that could transform both user's search and database words into a single canonical form that is known as Conflation is introduced. One of the well-known conflation algorithms, stemming, is only used to identify morphological variants. Stemming algorithms are language dependent. They have proven to be successful in reducing words with the same stem to a common form and are evidenced by the work of many researchers. Unfortunately, the conflation method is unable to conflate different words that possess the same meaning. These words can only be conflated by a thesaurus that can handle hierarchic, synonymic, and also morphological relationship. To create a thesaurus for a given subject an extensive manual and highly skilled so to solve this problem, another language dependent conflation method, thesaurus is used. It can build all types of relationships that exists between words. The information retrieval thesaurus typically contains a list of terms, where a term is either a single word or phrase. The relationships between them are also included to assist in coordinating indexing and retrieval. So it is found that the incorporation of stemming algorithm and thesaurus successfully increase the retrieved and relevant documents using Malay query words.

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

The study on information retrieval is on 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 sore, 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. There is two main type of research area in information retrieval that the main objective is to enable end-user to perform searching effectively and efficiently. One of those areas is the knowledge-based approach to information retrieval systems that uses expert systems techniques to encode the expertise possessed by a trained intermediary. The other research is focus on the development of algorithmic procedures, which allow the computer to undertake the functions of a trained intermediary. This approach is known