

UNIVERSITI TEKNOLOGI MARA

**INFORMATION RETRIEVAL SYSTEM (IRS)
PROCESS ON HANSARD DOCUMENT**

SURAYA BINTI SAAD

Master

December 2007

ABSTRACT

Retrieving relevant speeches from Hansard document is one of the current problems in our Parliament. This research discusses about Information Retrieval process before we can search information on Hansard document easily and precisely. This research develops an algorithm which will enhance the current system. One of the IR process is to split the document into subdocument to ease searching relevant information. The algorithm will be developed to split the document into few subdocuments by using keywords. This research emphasizes that understanding how a data structure is used makes it clear how it is organized before we build an index to ease understanding of the construction algorithm which is usually more complex. Therefore with a proper process and use the right keywords the relevant information could be searched easily and precisely.

ACKNOWLEDGEMENT

In the Name of Allah. The All Merciful, The Most Merciful
Praise be to Allah S.W.T. Most Gracious, Most Beneficent.

This project is culmination of pursuit, which involved the time, and effort of many people whom I wish to acknowledge.

I would like to express my most sincere appreciation and gratitude to my supervisor, Prof. Dr Zainab Binti Abu Bakar for her guidance, advice and moral support for this project development from the beginning to the end. Her constant encouragement and enthusiasm for new ideas were a source of great inspiration.

I am also thankful to all lecturers at Faculty of Information Technology and Quantitative Science (FTMSK) for their assistance. Also thank you to En Samin of Turbine Techniques Sdn Bhd for giving me the ideas on Hansard process in Parliament and current system information. His information is really helpful in this research.

Last but not least, an expression of gratitude for my beloved husband, Mohd Zaid Bin Abdul Rani who always assist and give moral support and never give up courage me during my study. His encouragement and patience has made this work possible. Also thank you to all my daughters Nur Syaheeda, Nur Syafeeqah, and Nur Syaheerah and my son Muhammad Adam Zulhilmi for their love, understanding and support throughout my study. Also to my friends Azman Ali and others who have contributed for this project. Without whose presence, encouragement and understanding this pursuit would not have been a success. Again thank so very much to all of you.

TABLE OF CONTENTS

	PAGES
TITLE PAGE	i
ABSTRACT	ii
CANDIDATE'S DECLARATION	
ACKNOWLEDGEMENT	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	ix
CHAPTER 1	INTRODUCTION
1.1	Background 1
1.2	What is Hansard 3
1.3	Producing Hansard 3
1.4	AIM 4
1.5	Objectives of the Research 4
1.6	Benefits 5
1.7	Scope of Research 6
1.8	Significance Study 6

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Methods of searching for information in documents, searching for documents themselves, searching for metadata which describe documents, or searching within databases, whether relational stand-alone databases or hyper textually-networked databases such as the World Wide Web is called Information Retrieval (IR). Data retrieval, document retrieval, information retrieval, and text retrieval, and each of these has its own bodies of literature, theory, praxis and technologies. IR is, like most nascent fields, interdisciplinary, based on computer science, mathematics, library science, information science, cognitive psychology, linguistics, statistics, physics[1].

Problem of information storage and retrieval has attracted increasing attention because enormous amounts of information to which accurate and speedy access is becoming ever more difficult. The effect is that relevant information gets ignored since it is never uncovered, which in turn leads to much duplication of work and effort. Computer's technology provides rapid and intelligent retrieval systems such as in libraries where many of which certainly have an information storage and retrieval problem. Tasks, such as cataloguing and general administration, have successfully been taken over by computers. However, the problem of effective retrieval remains largely unsolved. Furthermore, the IR has changed considerably in the last years with the expansion of the WWW (World Wide Web) and the advent of modern and inexpensive graphical user interfaces and mass storage devices (C.J. Van Rijsbergen, 1979).