University of Technology MARA

TEXT RETRIEVAL FOR MALAYSIAN PARLIAMENT DOCUMENT USING LATENT SEMANTIC INDEXING TECHNIQUE

By

NOR HIDAYAH BINTI ISHAK
2008287194

Thesis submitted in fulfillment of the requirements for
Bachelor of Science (Hons) in Information System Engineering
Faculty of Computer and Mathematical Sciences

May 2010

ABSTRACT

Information retrieval is a tool to perform searching an effectively and efficiently. Document retrieval or text retrieval is a branch of information retrieval where the information is stored in the form of text. According to Tengku Mohd, document retrieval in Information Retrieval Systems (IRS) is generally about understanding of information in the contents of documents (2005). The retrieval outcomes will be more effective if the system more understand the contents of documents (Tengku Mohd, 2005).The rapid growth of text document in Malay language such as parliamentary proceedings, newspaper articles, and academic publications can give harder to find relevant information. Therefore, this research is developed text retrieval prototype using Latent Semantic Indexing on transcribed Malaysian Parliament documents. By using Latent Semantic Indexing (LSI) technique is can speed up the process of searching and retrieving information from text documents. The Malaysian Parliament documents are maintained by House of Parliament. These documents are categorized into House of Commons and House of Lords.

ACKNOWLEDGEMENT

TABLE OF CONTENTS

CHAPTER 1

OVERVIEW

This chapter is discussing about the background and rationale for the study. Therefore, this chapter emphasis on the research background that significantly relates to the problem statement, research scope, research objectives and significant of the research.

1.1 Introduction

According to Baeza-Yates and Ribiero-Neto, the area of information retrieval has grown well in the past 20 years ago with primary goals of indexing text and searching for useful documents in a collection (1999). The main function of information retrieval systems is to provide the users with tools to perform searching effectively and efficiently. Nowadays, example researches in information retrieval are modeling, document classification and categorization, system architecture, user interfaces, data visualization, filtering, languages and so on (Baeza-Yates and Ribiero-Neto, 1999). According to Ren and Bracewell, information retrieval is (IR) is "the act of storing, searching, and retrieving information that match a user's request "(2009). However, according to Tengku Mohd, information retrieval can be defined as "the study of how to determine and retrieve from a corpus of stored information the portions which are relevant to particular information needs" (2005). According to Wiesman et al., the function of an information retrieval system is to make the users to satisfy their information needs (1997). This information can be categorized as a text, sound, images, video, and so on (Wiesman et al., 1997).

Text retrieval also known as text search engine is a key technology in information retrieval. Text search engines are tools for finding documents in a collection such as parliamentary proceedings, newspaper articles, academic

1