Relevance Feedback for Malay Hadith Search Engine using Support Vector Machine

BY

HAJAR IZZATI BT MOHD GHAZALLI

BACHELOR OF COMPUTER SCIENCE (Hons)

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR OF COMPUTER SCIENCE

FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES

UNIVERSITI TEKNOLOGI MARA

MAY 2011

Acknowledgement

Alhamdulillah, praises to Allah, Lord of Al-Mighty, for His Guidance and will, for the revelation of some of His knowledge for me in the successful to develop this project.

This project would also not be possible and successful without the help and support from the course co-ordinator, Dr. Riaza Mohd Rias and supervisor, Assoc. Prof. Dr. Nurazzah Abdul Rahman. I would like to acknowledge them for giving me the instructions, advices, motivation, support and guidance in developing this project.

Also, I would like to thank my mother and my siblings who have given a lot of support and motivation throughout the time.

Not forgetting to mention and extend thanks to all friends and others for their opinions, supports, suggestions and cooperation to do this research.

Thank you.

Abstract

Finding useful and relevance information is frequently a tiresome and difficult task. Therefore, Relevance Feedback is needed as it aims for more efficient result and more satisfaction from the users. Relevance feedback helps to improve the retrieval process by ranking the documents. By employing a relevance feedback technique in Mutiara Hadith Search Engine, it can facilitate the users to find related hadith based on their query(s). Furthermore, only related hadith will be displayed, which the users have chosen. This project adopts Support Vector Machine technique in Malay Hadith Search Engine and the effectiveness of the technique is evaluated using recall, precision and effectiveness. This project shows an increasing effectiveness when the technique is adopted.

Keywords: Information Retrieval (IR), Relevance Feedback (RF), Support Vector Machine (SVM), retrieval effectiveness

Table of Contents

CHAPTER 1

Introduction

1.1 Background

The rapid growth of online information and services on the net provides a rich knowledge resource. The World Wide Web (WWW) has become more popular for people to get information (Aljlayl and Frieder, 2001). Based on the 2004 Pew Internet Survey (Fallows, 2004), 92% of Internet users say the Internet is a good place to go for getting everyday information.

Finding useful information on the Web is frequently a tiresome and difficult task. For instance, to satisfy his information need, the user might navigate the space of Web links searching for information of interest. This might take a lot of time to filter the information. Therefore, information retrieval or IR was introduced.

IR involves finding some desired information in a store of information or a database (T. Meadow *et al.*, 2007). IR is fast becoming the main form of information access over taking the traditional database style searching. In IR, relevance feedback is a commonly accepted method of improving retrieval effectiveness (Rocchio, 1971; Salton, 1989; Salton and Buckley, 1990).

Relevance Feedback or RF (Burkley *et al.*, 1994) aims for more efficient retrieval method and more satisfaction from users. It is the most common