Publication Retrieval System for Bahagian Arkib & Muzium (BAM) UiTM using Vector Space Model

BY

NOR ADZLAN BIN JAMALUDIN

BACHELOR OF COMPUTER SCIENCE (HONS)

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

FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES

UNIVERSITI TEKNOLOGI MARA

MAY 2011

Acknowledgement

Praise to Allah for giving me the blessing, health and guidance to complete this proposal. I would like to express my gratitude to PM. Dr. Nursuriati Jamil who was willing to become my supervisor and Dr. Riaza Rias for being my lecturer and coordinator. Their assistance, guidance and suggestion have been extremely helpful in the completion of this proposal. In addition to this, I would also thank the other lecturers, my family and friends for the support that they have given me throughout the process of this project.

Abstract

Information retrieval and search engines have become an essential aspect in many information systems. In systems which employ large databases, information retrieval becomes more important as it is able to perform better than the standard query search depending on the model used. Currently, the Bahagian Arkib & Muzium (BAM) of UiTM does not have an information search engine for its publication and retrieves information on publication materials manually. The objectives of this project are to create a publication bulletin database for BAM, create a retrieval engine prototype based on the bulletin database and test the functionality of the prototype in retrieving the information on bulletins. A search engine prototype that is based on the vector space model is implemented for the project. This prototype will extract the data from the database and index the data to create an inverted file. The inverted file will then be used for the retrieval process by comparing it with the query submitted by the user to identify the relevant publication materials and displaying a sample of the document back to the user. The prototype is able to successfully retrieve data from the database. However, the retrieval method is limited to the effectiveness of the stemmer and stop-word removal being used and further research should be conducted in these areas.

Keywords: Information Retrieval, Search Engine, Vector Space

Table of Contents