UNIVERSITI TEKNOLOGI MARA

Semantic Analysis of Hadith for Topic Classification Using Latent Semantic Indexing (LSI)

Aiman Haziq Bin Ibrahim

Thesis submitted in fulfilment of the requirements of Bachelor of Computer Science (Hons.) College of Computing, Informatics and Mathematics

February 2024

ACKNOWLEDGEMENT

Alhamdulillah, praises and thank you to Allah because of His Almighty and His utmost blessings, I was able to complete this research within the designated timeline. I would like to extend my sincerest thanks to my supervisor, Roslan Sadjirin, for his guidance and support throughout the project journey. Without him, I would not have been able to complete this project with such success.

I would like to extend my heartfelt gratitude to my lecturer, Madam Ummu Fatihah Binti Mohd Bahrin, for her support and knowledge in the subject of CSP650 that have greatly assisted me in my research. Furthermore, my beloved parents also deserve a special mention for their unwavering support and encouragement during this project. I am forever grateful to them for their love and support.

Finally, I would like to express my appreciation to my dear friends and everyonewho has been a part of this journey and has provided me with their unwaveringsupport and encouragement. Without them, this project would not have been possible.

ABSTRACT

The aim of this project is to provide a framework utilizing Latent Semantic Indexing (LSI) to categorize topics in Hadith texts for semantic analysis. Islamic teachings place a high value on the hadith literature, which records the words and deeds of Prophet Muhammad (peace be upon him). To make it simple to access, retrieve, and comprehend pertinent information, Hadith writings must be effectively organized and categorized depending on their topics. The subjectivity, labor-intensive manual categorization, and insufficient capture of semantic links within texts are only a few of the drawbacks of the currently available approaches for Hadith topic classification. To address these challenges, LSI-based framework was proposed that leverages the latent semantic meaning in Hadith texts. LSI captures the underlying semantic relationships between words and enables more accurate topic classification. The research framework consists of six phases, including a preliminary study, requirement analysis, data finding, development, evaluation, and documentation. The data finding involves collecting and preprocessing reliable Hadith datasets. Development focuses on creating an information retrieval system using LSI. The evaluation assesses the system's performance through metrics like cosine similarity, precision, recall, and F1 Score. The experiment assessed the effectiveness of LSI by utilizing ten queries and relevant judgements, precision ranged from 5.4% to 100%, recall from 0% to 65%, yielding an average F1 Score of 19.4%. Finally, documentation encompasses writing a comprehensive report that includes background, methodology, findings, and conclusions.

TABLE OF CONTENTS