Universiti Teknologi MARA

Topic Based Search For Malay Translated Al-Quran Documents

Yuhaini Binti Yusop

Thesis submitted in fulfillment of the requirements for
Bachelor of Computer Science (Hons.)
Faculty of Computer and Mathematical Science

JANUARY 2014

ACKNOWLEDGEMENT

Bismillahirrahmanirrahim

In the name of Allah. The most Gracious and Most Merciful

First and foremost, I would like to praise Allah, because of His Almighty and His utmost blessings, I finally able to finish this research within the time duration given. This project is fulfilment of the requirements for Bachelor of Computer Science (Hons). I am thankful for the opportunity to conduct this research and for the strength that Allah S.A.W gave me throughout the process.

Secondly, special thanks to my supervisor Puan Norasiah Mohammad for her outstanding guidance and valuable advice, who took keen interest for all my work all over the year, both in the implementation part and writing this thesis. I appreciated her time for helping, support and encouraging me throughout the project is conducted.

Thirdly, thanks to my beloved family for always supporting and trusting me whatever I do especially my parent. Thanks for the love and kindness and always be there for me, they are always in my heart.

Last but not least, I would like to give my gratitude to my dearest friend for the collaboration and for everything that they have done for me during this research is carried out.

ABSTRACT

Document classification retrieval is a method of selecting the right document that contains useful information. One of the data corpuses that can be stored in the document is Al-Quran Malay translation collection. Al-Quran is a holy book and can be extremely important document to Muslims. It is used as a guidance to lead a righteous life. Most people read the Al-Quran along with its translation to understand the content or the meaning of the verses in each of the Surah (chapters). Each chapter describes information that can be referred to by Muslims with the help of the translation. However, it is not an easy task to search for information in the Malay translated Al-Quran since it contains many pages. Most of people find it hard to search the phrases manually in the Al-Quran, and it require a lot time consuming. Moreover, most of the searches are not related to what the user want. Therefore, the objective of this project is to develop Malay search query prototype for the Malay translated Al-Quran document where the user can search the information based on topic. Besides that, this project will also focus on the Malay text language, and the data to be used are the Malay translation of chapter Al-'Imran. Thus, to accomplish these, the technique of Latent Dirichet Allocation (LDA) is proposed to extract the topic in the document. The significance of this project is to ease the effort of people in finding specific topic in the Al-Quran and the related information from those topics. Moreover, it can also reduce the time to search for the word and to provide fast retrieval of data. The expected outcome of this project is Malay search query prototype for the Malay Translation Al-Quran Document to extract documents which has verses containing the topic.

Keywords— Document Classification, Malay Translated Al-Quran Document, Latent Dirichlet Allocation, Topic.

TABLE OF CONTENTS