UNIVERSITI TEKNOLOGI MARA

SYSTEM FOR RETRIEVING SIMILAR SENTENCES FROM MALAY TRANSLATION OF THE AL-QURAN

NUR FARHANA BT RASIP

Thesis submitted in fulfilment of the requirements for

Bachelor of Computer Science (Hons.)

Faculty of Computer and Mathematical Sciences

JULY 2019

ACKNOWLEDGEMENT

I would like to thank to Allah Almighty for giving me his blessing in completing my CSP650 project formulation for project successfully in the time that has been given.

I would like to take this opportunity to convey my appreciation to my supervisor Pn Haslizatul Fairuz Binti Mohamed Hanum for her guidance, encouragement, comments, ideas and support in order to make me completed this work better. Lasst but not least, I would like to thank to DR. Marina Ismail for her guidance throughout in completing this report proposal.

I also would like to thank you to my family for their support and encouragement to make me able to complete this proposal completely. May Allah S.W.T bless them.

Thank You.

ABSTRACT

Similar sentences can be computed by using Bigram Language Model. There are others researchers computed similar sentences but not using Malay documents. The problems are there are lack of study related to Malay documents for computing similar sentences. It will also produce huge amount of relevant documents by using frequent words and topic words. The similar sentences also were scattered in the Al-Quran. So, the user will have a hard time to search what they want manually. There are two process that need to be conducted to retrieve similar sentences. The first process was pre-processing the data and used Bigram Language Model to retrieve similar sentences. Bigram Language Model will counts all 2-word-long subsequence or bigram that appear on data and build probability distribution of bigram. As the result, the prototype will display to user the similar sentences that match with user's query. The precision and recall was high for every query. The benefits of this system was instead of a user manually searched for what they want in the Al-Quran by flipping through the translated documents, this system will help user to quickly enters a query word and retrieve all the match documents that are related to what they are looking for.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.0 Introduction

This chapter provides a general introduction of the project and the overview of this project. This chapter covers about the background of study, problems statement, objectives, scopes and significance of the research. This chapter is an important part for the understanding of this project.

1.1 Background of study

There are many Quranic verses that describe a concept or story in Quran. The relevant verses can be retrieved using keyword matching (Al-hagery, 2016). However, user will get huge amount of relevant verses as result from the search (Nor Diana Ahmad et al, 2017). The project will explore on a technique to refine element result extracted from retrieval system.

1.2 Problem Statement

Using keyword matching approach that matches a user query and keywords from documents, a search engine retrieved huge number of relevant documents retrieved results (Nor Diana Ahmad et al, 2017). The common global ranking approaches is to deliver relevant results based on frequent keywords that describe a theme or topic in the documents. However, for some type of documents, not all theme/topic can be represented as frequent keywords.

Based on (Husin, 2017) says that studies in unstructured text especially in Malay language are quite not popular among researchers due to lack of availability extraction tools to extract relevant information and studies in Malay