

Universiti Teknologi MARA

**E-History Malaysian Secondary School
Textbook Using TF-IDF Algorithm and
Text Visualization**

Nur Hafizah Mohd Ridzuan

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

July 2020

SUPERVISOR APPROVAL

E-HISTORY MALAYSIAN SECONDARY SCHOOL TEXTBOOK USING TF-IDF ALGORITHM AND TEXT VISUALIZATION

By

NUR HAFIZAH BINTI MOHD RIDZUAN
2017412246

This thesis was prepared under the supervision of the project supervisor, Nurulhuda Binti Zainuddin. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfilment of the requirements for the degree of Bachelor of Computer Science (Hons.).

Approved by



.....
Nurulhuda Zainuddin
Project Supervisor

JULY 10, 2020

STUDENT DECLARATION

I certify this thesis and the project to which it refers is the product of my own work and that any idea of quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

Hafizah

.....
NUR HAFIZAH BINTI MOHD RIDZUAN
2017412246

JULY 10, 2020

ABSTRACT

History is one of the school subjects in Malaysian schools and known as a compulsory subject that is required to pass the SPM examination. One of the main learning methods in Malaysian education is by using textbooks as it is a universal formal school learning material and a dominant resource of learning. However, the use of textbooks for historical subjects is difficult due to the highly textual content of the textbook itself and the heavy textbook burden for the user. Hence, E-History is project that proposed to help in overcome this problem. The objective of the project is to design and develop an E-History Malaysian secondary school textbook system using Term Frequency-Inverse Document Frequency (TF-IDF) algorithm with text visualization and also to test the functionality and usability of the system through a web-based system. TF-IDF Algorithm technique is performed to develop system by calculating the most relevant pages in the textbook based on keyword that entered by user's query. Text visualization technique will generate word cloud visualization that used to visualize each chapter by visualize the most relevant based on TF-IDF score. The expected result, user get the visualization of word cloud that present the size of word based on frequencies of most appeared on the chapter. The finding of this system is believed to be helpful as it is chip in to the education materials. Results show that all functions in the system is fully functioned and well-integrated plus most of the users provide good and promising feedbacks during the usability testing. As the future works, the system can be improved by applied the system using mobile apps application to make the system can have easier access and implementation of another technique such as semantic-based searching and the usage of stemming algorithm would be considered to produce more meaningful results.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	x
LIST OF TABLES	xiii

CHAPTER ONE: INTRODUCTION

1.1	Background of Study	1
1.2	Problem Statement	3
1.3	Project Objective	4
1.4	Project Scope	5
1.5	Significance of the Study	5

CHAPTER TWO: LITERATURE REVIEW

2.1	Education of Malaysia	6
2.2	Textbook	8
2.2.1	References Book	8
2.2.2	E-Textbook	9
2.2.3	Open Textbook	9
2.3	Malaysian Education Textbook	10
2.4	Information Retrieval	11
2.4.1	TF – IDF Technique	11
2.4.2	Boolean Model	13
2.4.3	Vector Space Model	14