

UNIVERSITI TEKNOLOGI MARA

**ARTICLE RECOMMENDATION SYSTEM USING
CONTENT-BASED FILTERING**

MUHAMMAD HAZRUL AFIQ BIN KAMARUDDIN

BACHELOR OF COMPUTER SCIENCE (Hons.)

JANUARY 2024

ACKNOWLEDGEMENT

My sincere appreciation goes out to our wonderful lecturer, Madam Ummu Fatimah Mohd Bahrin, for her invaluable guidance and support during this final year project. Her solid commitment to professionalism and her passion for sharing knowledge have greatly influenced the route this project has taken.

I would especially like to express my gratitude to my supervisor, Pn Nurhilyana Anuar, whose advice, and helpful criticism have been extremely helpful in improving the project's quality. Her guidance has helped me to reach this far and be able to complete the final year project.

Lastly, I also want to thank my family and fellow friends for their guidance and moral support during this journey. They also have their own contribution, and this project was successfully completed thanks to them.

ABSTRACT

The challenges in the education industry include the time-consuming process of manually searching for relevant research articles, which reduces productivity and negatively impacts academic performance. Furthermore, the growing issue of information overload and anxiety among students and researchers raises the risk of burnout and decreases overall academic performance. In this study, article recommendation system using content-based filtering was designed and developed to address the challenges. The algorithm used is able to generate relevant article recommendations based on the content of the article. The algorithm consists of two components which are Term Frequency – Inverse Document Frequency (TF-IDF) and Cosine Similarity. TF-IDF calculates the weightage of user query and each keyword in each article. Vectors containing weightage values for both user query and articles in dataset will be calculated using Cosine Similarity to obtain similarity value. Articles recommendation will be generated after sorting and filtering based on threshold value. The result was evaluated using confusion matrix and evaluation metrics such accuracy, precision, recall and F1 score. The article recommendation system is able to achieve up to 99% accuracy, 86% precision, 76% recall and F1 score of 0.8 where the threshold value is 0.1. Overall, the project is successful as it is able to generate relevant articles accurately.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	II
STUDENT DECLARATION	III
ACKNOWLEDGEMENT	IV
ABSTRACT	V
TABLE OF CONTENTS	VI
LIST OF FIGURES	IX
LIST OF TABLES	XI

CHAPTER 1: INTRODUCTION

1.1	BACKGROUND OF STUDY	1
1.2	PROBLEM STATEMENT	2
1.3	OBJECTIVES	4
1.4	PROJECT SCOPE	4
1.5	PROJECT SIGNIFICANCE	5
1.6	OVERVIEW OF RESEARCH FRAMEWORK	6
1.7	CONCLUSION	6

CHAPTER 2: LITERATURE REVIEW

2.1	RECOMMENDATION SYSTEM	8
2.1.1	Introduction to Recommendation System	8
2.1.2	Type of Recommendation System	9
2.2	ARTICLE RECOMMENDATION SYSTEM	12

2.3	MACHINE LEARNING	13
2.3.1	Supervised Learning	13
2.3.2	Unsupervised Learning	14
2.3.3	Semi-Supervised Learning	14
2.3.4	Reinforcement Learning	15
2.4	CONTENT-BASED FILTERING	15
2.4.1	Text Extraction	15
2.4.2	Text Similarity	18
2.4.3	Implementation of Content-Based Filtering in Various Problems	21
2.5	SIMILAR WORKS	31
2.6	THE IMPLICATIONS OF LITERATURE REVIEW	46
2.7	CONCLUSION	47

CHAPTER 3: METHODOLOGY

3.1	OVERVIEW OF RESEARCH FRAMEWORK METHODOLOGY	49
3.1.1	Detailed Research Framework	49
3.2	PRELIMINARY STUDY	54
3.2.1	Knowledge Acquisition	54
3.2.2	Literature Review	55
3.3	DATA COLLECTION AND PREPARATION	55
3.3.1	Data Collection	56
3.3.2	Data Preparation	57
3.4	DESIGN AND IMPLEMENTATION	58
3.4.1	Software Development Life Cycle (SDLC)	58
3.4.2	Prototype Architecture	59
3.4.3	Flowchart	60
3.4.4	Pseudocode	62
3.4.5	Interface Design	63
3.5	SYSTEM DEVELOPMENT	64
3.5.1	Development	64
3.5.2	Hardware and Software Recommendation	65