

UNIVERSITI TEKNOLOGI MARA

**TOWARDS DEVELOPING A WEB APPLICATION
WITH TEXT FILTERING SYSTEM FOR FTMSK:
A CASE STUDY FOR ELECTRONIC EXCHANGE BOARD**

SUMARDI BIN SHUKOR

Bsc. (Hons.) INFORMATION SYSTEM ENGINEERING

NOVEMBER 2006

DECLARATION

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

NOVEMBER 2, 2006

SUMARDI BIN SHUKOR
2004106722

TABLE OF CONTENTS

| | |
|---|-------------|
| ACKNOWLEDGEMENT | iii |
| LIST OF TABLES | viii |
| LIST OF FIGURES | ix |
| LIST OF ABBREVIATIONS | x |
| ABSTRACT | xi |
| CHAPTER 1 : INTRODUCTION | |
| 1.1 Background | 1 |
| 1.2 Problem Statement | 2 |
| 1.3 Objectives Of The Project | 4 |
| 1.4 Scope | 4 |
| 1.5 Significance Of The Project | 5 |
| 1.6 Limitation Of The Project | 5 |
| 1.7 Overview Of The Project | 6 |
| CHAPTER 2 : LITERATURE REVIEW | |
| 2.1 Introduction | 7 |
| 2.2 The Definition Of Filtering | 7 |
| 2.3 Review On Information Filtering | 8 |
| 2.4 Electronic Junk | 8 |
| 2.5 Existed Algorithm of Information Filtering | 10 |

| | | |
|-----|--|----|
| 2.6 | The Issues that related to Electronic Exchange Board System | 15 |
| 2.7 | Methods That Have Been Used To Mask Offensive Words | 16 |
| 2.8 | Regular Expression for Textual Filtering Design | 19 |
| 2.9 | Apache Web Server For The Development Of Filtering Module | 21 |

CHAPTER 3 : PROJECT APPROACH AND METHODOLOGY

| | | |
|-------|---|----|
| 3.1 | Introduction | 23 |
| 3.2 | Project Assessment and Project Planning | 23 |
| 3.2.1 | Problem Assessment | 23 |
| 3.2.2 | Project Planning | 23 |
| 3.3 | Knowledge Acquisition | 24 |
| 3.3.1 | Observation of Online Journals and Articles | 24 |
| 3.4 | Data Analysis | 24 |
| 3.4.1 | Knowledge Abstraction for Primary Sources | 24 |
| 3.4.2 | Knowledge Abstraction for Secondary Sources | 25 |
| 3.5 | Design | 25 |
| 3.6 | Development | 25 |

CHAPTER 4 : FILTERING ALGORITHM SIMULATOR DEVELOPMENT

| | | |
|-------|-------------------------------|----|
| 4.1 | Introduction | 26 |
| 4.2 | Overview | 26 |
| 4.3 | Filtering Module (mod_censor) | 26 |
| 4.3.1 | Process Flow | 27 |
| 4.3.2 | Request Processing | 28 |
| 4.3.3 | Request Processing Phase | 29 |
| 4.3.4 | The Data Axis And Filter | 30 |
| 4.3.5 | Filtering Module | 31 |

ABSTRACT

The Internet is the most disembodied media as it is the most removed from face-to-face-communication. In Internet communication we do not have the limitations of space. There are no borders for circulation and there is unlimited memory space for content. This dimension of disembodiment constitutes the globalisation of the medium. On the other hand, Internet-based communication has increased opportunities to secure information in a manner that has not been available up to now. This project purpose is to suggest the usage of filtering system as a mean of information filtering in the Electronic Exchange Board for FTMSK. The traditional filtering mechanism in the Electronic Exchange Board might not be able to distinguish the offensive words that have been masked. Hence, this study is conducted to identify potential techniques can be used to mask the offensive words and a new textual filtering algorithm will be developed to resolve this problems. Most of the identified masking techniques can be detected by this new filtering algorithm. However, it can never successful entirely because people will always find new ways to mask the offensive words. Therefore, there should be an authorized person whose job is to monitor and determine which postings can be allowed and not in Electronic Exchange Board.